Identification of breast cancer prognostic modules based on weighted protein-protein interaction networks

WAN LI[1*], XUE BAI[1*], ERQIANG HU[1*], HAO HUANG[1], YIRAN LI[1],
YUEHAN HE[1], JUNJIE LV[1], LINA CHEN[1] and WEIMING HE[2]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086;
[2]Institute of Opto-electronics, Harbin Institute of Technology, Harbin, Heilongjiang 150001, P.R. China

Abstract. Breast cancer is one of the leading causes of mortality in females. A number of prognostic markers have been identified, including single genes, multi-gene signatures and network modules; however, the robustness of these prognostic markers is insufficient. Thus, the present study proposed a more robust method to identify breast cancer prognostic modules based on weighted protein-protein interaction networks, by integrating four sets of disease-associated expression profiles. Three identified prognostic modules were closely associated with prognosis-associated functions and survival time, as determined by Cox regression and Kaplan-Meier survival analyses. The robustness of these modules was verified with an independent profile from another platform. Genes from these modules may be useful as breast cancer prognostic markers. The prognostic modules could be used to determine the prognoses of patients with breast cancer and characterize patient recovery.

Introduction

Breast cancer is one of the leading causes of female mortality (1-3). Prognosis reflects the outlook and chance of recovery from breast cancer. It is therefore critical to acquire prognostic information that represents the physical condition of the patients. Understanding prognostic information assists in increasing the survival rate and prolongs the life expectancy of patients (4).

With the advent of microarray and next-generation sequencing technology, an increased volume of genomic data is available, including data relevant to breast cancer prognosis. In recent decades, a number of prognostic marker identification methods based on a single gene have been proposed (5). However, the detection of the expression of a single gene does not effectively characterize the expression of a gene cluster composed of tens or hundreds of genes. It has been demonstrated that prognostic information from multi-gene signatures, including network modules, can reduce the heterogeneity of diseases with greater accuracy than data of a single gene (6). For example, a classifying system for breast cancer profiles was constructed based on 70 genes in a study by van de Vijver *et al* (7). A treatment response prediction method for patients with breast cancer was developed based on 64 genes by Pawitan *et al* (8). The 21-gene signature has now been intensively studied and is widely used in clinical decisions regarding breast cancer (9).

Prognostic marker identification methods have the problem of low robustness (10), and prognostic information obtained from certain patients may not be applicable to other patients (11). Thus, the present study aimed to develop a method to identify robust prognostic modules based on four breast cancer expression profiles and protein-protein interaction networks (PPINs). Experimental detection based on these prognostic modules could aid clinicians in the diagnosis of patients who are at risk of developing malignant disease, and allow the provision of prevention and treatment for these patients as early as possible.

Materials and methods

Expression profile data. In total, 5 human breast cancer expression profiles, with survival time data, were downloaded from the Gene Expression Omnibus (GEO) database (12). GSE1456 (159 samples) (8), GSE2034 (286 samples) (13), GSE3494 (251 samples) (14) and GSE4922 (249 samples) (15) were all originally created using the Affymetrix, Inc., Hu133A and Hu133B platform (Affymetrix, Santa Clara, CA, USA). The GSE1456 dataset was selected as a training set, with the others were used as validation sets, to identify prognostic modules. GSE9893 (155 samples) (16), a profile created with MLRG Human 21K V12.0 (IGMM, Montpellier, France), was used to assess the validity and robustness of the prognostic modules.

Correspondence to: Professor Lina Chen, College of Bioinformatics Science and Technology, Harbin Medical University, 157 Baojian Road, Harbin, Heilongjiang 150086, P.R. China
E-mail: chenlina@ems.hrbmu.edu.cn

*Contributed equally

The protein interaction data was obtained from the Human Protein Reference Database (17). Weighted PPINs were constructed using Pearson correlation coefficients (PCCs) as weights between interacting proteins X and Y using expression values from the expression profiles, as follows:

$$PCC_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

n represents the number of samples and X_i and Y_i are the expression values of proteins X and Y in the ith sample, respectively. $P<0.05$ was considered to represent a statistically significant difference.

Identification of candidate prognostic modules. The network modules were mined from the PCC-weighted PPINs using clusterONE software, version 1.0 (18), which detected modules from weighted networks. Genes in each module were connected topologically and associated with expression.

Supervised principal component analysis (19) was applied to obtain candidate prognostic modules, taking the associated survival time and status (succumbed to disease or alive) into account. The mean expression value for each network module was calculated according to the expression profiles. The modules were screened against matrices composed of the mean expression values of network modules. The optimal feature threshold in supervised principal components was estimated by 10-fold cross-validation. Candidate prognostic modules were determined according to the optimal threshold.

Functional and survival analysis of candidate prognostic modules. Functional and pathway enrichment analyses were conducted to assess candidate prognostic modules, using functions from the Gene Ontology (GO) database (20) and pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (21). The significance threshold was set as $P<0.05$.

A univariate Cox regression analysis was performed to evaluate the association between survival time and the mean expression level of genes in candidate prognostic modules by statistical significance. Survival risks associated with candidate prognostic modules were evaluated by age-weighted hazard ratios.

Kaplan-Meier survival analysis was performed to assess the significance of modules, classifying patients according to the mean expression value of the module. Patients were separated into two classes, 'high' and 'low' expression, depending on whether their expression value was higher or lower than the mean of the module. Kaplan-Meier survival analysis was performed according to the survival time of patients in each class, to assess the significance of the effect of each module on survival time. The difference between the survival times of the two classes was compared with a log-rank test (22). $P<0.05$ was considered to represent a statistically significant difference.

A hypergeometric test was performed on each module pair in the network to identify modules that overlapped with each other more than expected by chance.

$$P = \sum_{o=b}^{\min(M,B)} \frac{\binom{M}{o}\binom{N-M}{B-o}}{\binom{N}{B}}$$

where N is the total number of genes in two modules, M is the number of genes in one module, B is the number of genes in the other module, b is the number of common genes. A cut-off value of 0.05 was set to screen out significantly overlapped modules.

Identification of prognostic modules. Four sets of candidate prognostic modules were identified for the four Affymetrix expression profiles. To ensure the robustness of prognostic modules, 'overlapping', i.e. recurring, genes from the four sets of candidate prognostic modules were extracted. Cox regression, Kaplan-Meier survival and functional enrichment analyses using GO and KEGG were then conducted to screen for significant modules, which were defined as breast cancer prognostic modules.

Results

Candidate prognostic modules. PCCs were calculated for expression values from the expression profile GSE1456. A weighted PPIN was constructed based on pairs with PCC>0.01 and $P<0.05$. ClusterONE identified 28 network modules that contained >4 genes, which were used for further analysis. These modules were designated as 'Module 1-28'.

There were a large number of overlapping genes in modules 1-28. A hypergeometric test was performed to assess the significance of overlapping in the modules. Modules overlapping significantly were merged, yielding 18 merged modules, designated as 'oModule 1-18'.

The optimal feature threshold in supervised principal components, estimated by 10-fold cross-validation, was 1.8. A total of 12 candidate prognostic modules (Table I) and 6 candidate prognostic merged modules (Table II) containing 46 and 48 genes, respectively, were produced according to the optimal threshold. All candidate prognostic merged modules were produced from the union of candidate prognostic modules, with the exception of Module 9.

Functional and survival analysis of candidate prognostic modules. Functional enrichment analysis was performed for the candidate prognostic merged modules. It was identified that 6 candidate prognostic merged modules (oModules 1, 2, 3, 4, 6 and 9) were enriched for GO functions or KEGG pathways, including cell cycle (a hallmark of malignancy) and DNA replication ($P<0.05$). For candidate prognostic modules that were not enriched in any function, e.g., Module 4, 7 and 18, the corresponding merged module, e.g., oModule 2, may be enriched in certain functions, including the cell cycle (23). Therefore, candidate prognostic merged modules were more significant at the functional level.

Candidate prognostic modules and merged modules were compared by hazard ratios and P-values calculated with univariate Cox regression analysis. P-values from Kaplan-Meier survival analysis were additionally computed to compare these types of modules. All modules were significantly associated with survival time (Cox, $P<0.0002$), and could significantly

Table I. Hazard ratio and significance of 12 candidate prognostic modules as assessed by Cox regression and Kaplan-Meier survival analyses.

Module	Genes, n	HR	Cox	Kaplan-Meier
1	15	3.95	7.8×10^{-5}	1.9×10^{-2}
4	5	3.7	2.2×10^{-4}	8.0×10^{-4}
7	5	3.72	2.0×10^{-4}	3.2×10^{-3}
10	6	3.91	9.4×10^{-5}	7.6×10^{-6}
13	6	3.73	1.9×10^{-4}	1.1×10^{-5}
15	5	4.06	4.8×10^{-5}	9.0×10^{-5}
16	5	3.81	1.4×10^{-4}	4.5×10^{-5}
17	5	3.82	1.3×10^{-4}	2.5×10^{-5}
18	5	3.63	2.8×10^{-4}	2.9×10^{-3}
19	6	3.91	9.2×10^{-5}	2.2×10^{-6}
20	5	3.77	1.6×10^{-4}	9.9×10^{-6}
27	5	4.07	4.8×10^{-5}	5.5×10^{-5}

HR, hazard ratio.

Table II. HR and significance of 6 candidate merged prognostic modules by Cox regression and Kaplan-Meier survival analyses.

olapModule	Genes, n	HR	Cox	Kaplan-Meier
1	17	4.05	5.2×10^{-5}	1.9×10^{-2}
2	7	3.76	1.7×10^{-4}	6.3×10^{-6}
3	7	3.86	1.1×10^{-4}	2.8×10^{-7}
4	6	3.91	9.4×10^{-5}	3.1×10^{-5}
6	6	3.87	1.1×10^{-4}	4.6×10^{-6}
9	5	4.07	4.8×10^{-5}	5.5×10^{-5}

HR, hazard ratio.

classify patients according to the mean expression values of genes in the module (Kaplan-Meier, P<0.02). Candidate prognostic merged modules were thus identified as more significant than the corresponding candidate prognostic modules with regard to analyzing survival time (Tables I and II).

The results indicated that 6 candidate prognostic merged modules were better than the candidate prognostic modules, as they represented candidate prognostic modules with improved functional enrichment, lower Cox hazard ratios and better classifying performance for survival time.

To verify the effectiveness of the 6 candidate prognostic merged modules identified based on the analysis of profile GSE1456, univariate Cox regression analysis and Kaplan-Meier survival analysis were performed for the modules using expression data from the other three Affymetrix profiles, GSE2034, GSE3494 and GSE4922. The 6 candidate prognostic merged modules were identified as significantly associated with survival time for the 3 expression profiles (P<0.05); the association with GSE3494 and GSE4922 exhibited particular significance (both P<0.01). Thus, the candidate prognostic merged modules were verified with other profiles from the same platform.

Breast cancer prognostic modules. The same process as for GSE1456 was performed with the profiles GSE2034, GSE3494 and GSE4922 to generate three sets of candidate prognostic merged modules. As overlapping genes represented the core characterization of the four sets of expression profiles, they were selected for their significance with a hypergemetic test, and five prognostic overlapping modules were obtained and designated as 'olapModule 1-5'.

To screen for significant modules that correlated with survival time from these prognostic merged modules, univariate Cox regression and Kaplan-Meier survival analyses were performed (Table III). OlapModules 1, 2 and 4 were selected as they were significantly associated with the survival time of patients in the two analyses (P<0.05). These modules were defined as breast cancer prognostic modules, and may be suitable for use as prognostic markers for patients with breast cancer.

The effectiveness of the 3 breast cancer prognostic modules was assessed by Kaplan-Meier survival analysis for all patients, integrating all four Affymetrix expression profiles (Fig. 1). This analysis demonstrated that the breast cancer prognostic modules could distinguish likely patient survival time based on their mean expression values, with statistical significance, when all Affymetrix expression profiles were integrated (P<0.0001).

The functional association between the prognostic modules and breast cancer prognosis was assessed by functional enrichment analysis. Genes in olapModule 1 were enriched in GO functions and KEGG pathways including DNA replication, nucleoside binding and cell division (P<0.05). All 3 prognostic modules were enriched for the 'cell cycle' GO function and the 'cell cycle' KEGG pathway (P<0.05). The majority of genes in olapModule 1 were associated with the G1 and S phase, whereas the genes of olapModule 2 and 4 were associated with the M and G2 phase, respectively (Fig. 2). OlapModule 1 was enriched for DNA replication and nucleotide binding, while olapModule 2 was enriched in cell division. The deregulation of cell cycle progression is one of the causes of breast cancer (22). It was previously demonstrated that cell cycle abnormalities in G1/S or G2/M phases are frequently observed in breast cancer (24-26). For example, deregulation of cell cycle events leads to uncontrolled cell proliferation, and cell proliferation dysfunction is the main cause of cancer, and nuclear division is significantly associated with breast cancer (27-29). A number of studies have implicated the modulation of the cell cycle and cell cycle regulatory proteins as markers for breast cancer prognosis (30,31). Therefore, the expression genes from these prognostic modules may reflect prognostic information, potentially providing guidance in the treatment of breast cancer.

A total of 23 genes were contained in the 3 prognostic modules. All these genes have been verified to be associated with breast cancer in previous studies, including 13 that have been verified to be breast cancer prognostic markers and 6 which are prognostic markers for other types of cancer (Table IV). The genes minichromosome maintenance complex component 2-7 (MCM2-7) in olapModule 1 form the mini-chromosomal maintenance protein complex, an important structure in

Table III. Assessment of the prognostic significance of olapModules for sets of expression profiles, as assessed by Cox regression and Kaplan-Meier survival analyses.

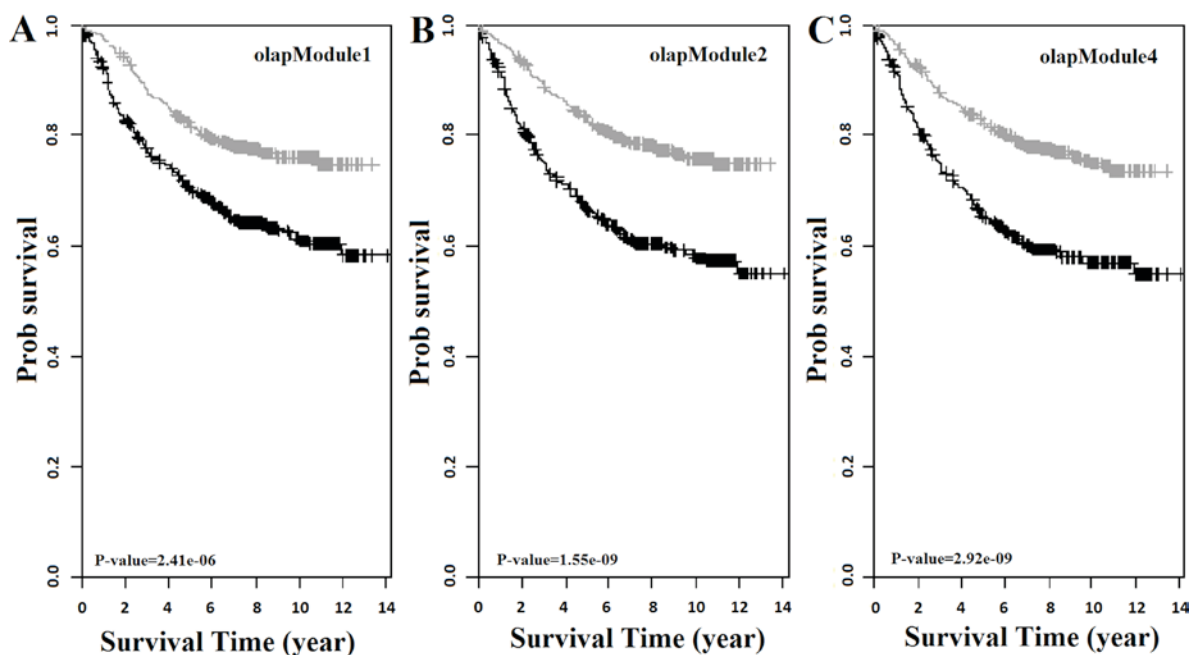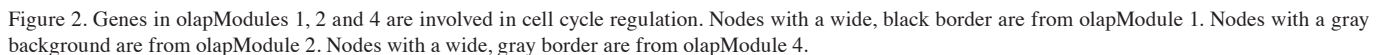| Profile set	olapModule	HR	Cox	Kaplan-Meier
GSE1456 (training set)	1	4.0	6.10×10^{-5}	6.40×10^{-4}
	2	3.9	6.90×10^{-5}	9.10×10^{-5}
	3	3.8	1.10×10^{-4}	4.60×10^{-6}
	4	3.7	2.20×10^{-4}	8.00×10^{-4}
	5	-1.9	5.10×10^{-2}	3.20×10^{-1}
GSE2034	1	2.2	2.20×10^{-3}	2.00×10^{-2}
	2	3.4	5.20×10^{-4}	2.10×10^{-3}
	3	1.8	6.40×10^{-2}	5.50×10^{-2}
	4	3.3	9.70×10^{-4}	1.50×10^{-2}
	5	-1.4	1.40×10^{-1}	6.40×10^{-1}
GSE3493	1	3.2	1.30×10^{-3}	1.00×10^{-2}
	2	3.5	4.40×10^{-4}	1.40×10^{-3}
	3	3.5	3.30×10^{-4}	3.10×10^{-2}
	4	3.8	1.20×10^{-4}	1.10×10^{-4}
	5	-2.9	3.40×10^{-3}	4.20×10^{-2}
GSE4922	1	3.6	2.60×10^{-4}	4.20×10^{-4}
	2	4.0	4.90×10^{-5}	5.40×10^{-4}
	3	3.9	8.70×10^{-5}	1.20×10^{-3}
	4	4.2	2.50×10^{-5}	3.00×10^{-4}
	5	-2.8	3.80×10^{-3}	1.80×10^{-2}

HR, hazard ratio.

Figure 1. The Kaplan-Meier survival curve for all patients in four sets of expression profiles, based on three breast cancer prognosis modules: (A) olapModule 1, (B) olapModule 2 and (C) olapModule 4. Gray and black lines indicate the survival time for patients with low- and high-expression, respectively.

DNA replication and extension. These genes could reflect the prognosis of patients and have been used as effective targets for cancer therapy (32-34) origin recognition complex subunit 1 (ORC1) and ORC6 in olapModule 1 are components of the DNA replication initiation complex, and have been used as breast cancer prognostic markers (35,36). The genes cell division cycle 20 (CDC20), ribonucleotide reductase regulatory subunit M2 (RRM2), cyclin dependent kinase inhibitor 3

Table IV. Literature verification for genes included in prognostic modules.

olapModule	Breast cancer marker	Cancer marker	Breast cancer-associated
1	MCM4 (25628920)	DBF4 (24287290)	MCM10 (25046975)
	MCM7 (19360293)	CDC7 (20724597)	CDC45 (17887956)
	CCNB2 (23282137)	MCM6 (21478909)	ORC6 (22333897)
	MCM2 (16278669)	MCM5 (20334671)	ORC1 (24097061)
	MCM3 (22699783)		
2	PTTG1 (23704896)	PKMYT1 (21042722)	
	MAD2L1 (25012665)	KIF11 (25193695)	
	CDC20 (17388661)		
	BUB1B (20569502)		
	CENPE (19470724)		
4	RRM2 (25016594)		
	CDK1 (25218592)		
	CDKN3 (20569502)		

Each number in brackets is a PubMed identifier to describe the marker or association status of the gene.

Figure 2. Genes in olapModules 1, 2 and 4 are involved in cell cycle regulation. Nodes with a wide, black border are from olapModule 1. Nodes with a gray background are from olapModule 2. Nodes with a wide, gray border are from olapModule 4.

(CDKN3) and MAD2 mitotic arrest deficient-like 1 (yeast) (MAD2L1) in olapModule 2 and 4 serve important roles in the regulation of cell proliferation, the cell cycle and chromosome segregation (37), whose abnormalities could cause breast

cancer. Genes in these modules that were not verified as associated with breast cancer prognosis could be novel prognostic markers for breast cancer.

Discussion

In the present study, a method has been proposed to identify robust breast cancer prognostic modules of multiple genes, based on weighted PPINs. Three prognostic modules were identified from the analysis of four sets of breast cancer-associated expression profiles using GO and KEGG functional enrichment, Cox regression and Kaplan-Meier survival analyses. These prognostic modules were significantly enriched in various prognosis-associated terms, including terms associated with the regulation of the cell cycle, which demonstrated the effectiveness of these modules from a functional perspective. The breast cancer prognostic modules are able to distinguish between patients based on their expression profiles not only from a single profile but also from a set of profiles integrating four sets, which demonstrated the robustness of the identified prognostic modules.

To further verify the robustness of prognostic modules, a set of expression profiles created with an alternative platform, GPL5049, was analyzed. Cox regression and Kaplan-Meier survival analyses were performed. A total of 3 prognostic modules were significantly associated with survival time for patients in this independent profile, which demonstrated the robustness of the 3 breast cancer prognostic modules.

Weighted PPINs were constructed based on PCCs by evaluating the correlation between protein pairs and their corresponding P-values. To confirm the advantages of utilizing weighted PPINs, unweighted PPINs were constructed to perform the same analysis. A total of 11 modules were detected using clusterONE and supervised principal component analysis, which were designated as 'nModule1-11'. No genes were overlapped between nModule 10-11, while a number of genes overlapped between the breast cancer prognostic modules. A total of 12 genes were common between nModule1 and olapModule 1, and there were 3 genes in common between nModule 4 and olapModule 2. Functional enrichment analysis was conducted for these 11 modules. Only nModules 1-4 were significantly enriched in biological functions, whereas only nModule1 and 4 were significantly enriched in the same functions as the breast cancer prognostic modules. In nModule 1, the genes were enriched in the same functions, including the cell cycle. The robustness of nModule 4 was not verified by the independent profile from another platform, as analyzed by Cox regression and Kaplan-Meier survival analyses (P>0.100). In addition, a greater number of genes in nModule 1 and 4 were not enriched in these functions, which indicated that olapModules 1 and 2 were more robust than nModule 1 and 4. Thus, modules based on unweighted PPINs were not effective as prognostic modules for breast cancer.

Breast cancer prognostic modules from the present study were compared with cancer prognostic signatures identified from another weighted network module-based method proposed by Wu and Stein (38). In the study by Wu and Stein (38), a single module was significant for Cox regression analysis across multiple datasets and was used as a prognostic signature.

Table V. HR and significance of the 21-gene signature using Cox analysis and Kaplan-Meier survival analysis.

Dataset	HR	Cox	Kaplan-Meier
GSE1456	3.9	3.6×10^{-2}	2.3×10^{-2}
GSE2034	1.0	4.5×10^{-1}	7.1×10^{-1}
GSE3494	2.2	2.3×10^{-1}	8.0×10^{-1}
GSE4922	5.2	1.1×10^{-3}	2.2×10^{-3}

HR, hazard ratio.

Kaplan-Meier survival analysis was subsequently employed. The prognostic signatures in Wu and Stein and in the present study were significant (P<0.05). However in the present study, the genes in these modules were different from the study by Wu and Stein. In the present study, all 23 genes in the modules have been verified to be associated with breast cancer in previous studies, 13 of which have been verified to be breast cancer prognostic markers (Table IV).

The widely used 21-gene signature (39) was assessed by Cox regression analysis and Kaplan-Meier survival analysis for the aforementioned expression profiles. The results for only 2 of 4 profile sets, GSE1456 and GSE4922, were significant (Table V), which indicated that the robustness of the 21-gene signature was not sufficient for certain patients.

By contrast, the robustness of the prognostic modules from the present study means that genes in these modules may be suited to become novel breast cancer prognostic markers. From a total of 23 genes contained in the 3 breast cancer prognostic modules, which have been previously verified as associated with breast cancer, 13 have been identified as prognostic markers for breast cancer, including ORC1, ORC6, CDC20, RRM2, CDKN3 and MAD2L1 (35-37) (Table IV). The remaining 10 genes may be novel breast cancer prognostic markers. Cox regression analysis and Kaplan-Meier survival analysis were performed for these genes. The majority of these genes were significantly associated with the survival time of patients (P<0.05) for four profiles, while particular genes that have been previously verified to be cancer prognostic markers, including ORC6, MCM3 and MCM7 (32-34,36), were not prognostic for certain expression profiles. These results demonstrated the advantage of using prognosis information from modules compared with using single genes.

In summary, the present study proposed an effective method to identify robust breast cancer prognostic modules. The proposed prognostic modules could be used to determine the prognosis of patients with breast cancer and characterize patient recovery. Genes in these modules may additionally act as breast cancer prognostic markers. This method may provide a novel perspective for the research of other types of cancer.

Acknowledgements

References

1. Fitzgerald SP: Breast-cancer screening. N Engl J Med 366: 191; author reply 191-192, 2012.
2. Hanahan D and Weinberg RA: Hallmarks of cancer: The next generation. Cell 144: 646-674, 2011.
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A: Global cancer statistics, 2012. CA Cancer J Clin 65: 87-108, 2015.
4. Colzani E, Liljegren A, Johansson AL, Adolfsson J, Hellborg H, Hall PF and Czene K: Prognosis of patients with breast cancer: Causes of death and effects of time since diagnosis, age, and tumor characteristics. J Clin Oncol 29: 4014-4021, 2011.
5. Staiger C, Cadot S, Györffy B, Wessels LF and Klau GW: Current composite-feature classification methods do not outperform simple single-genes classifiers in breast cancer prognosis. Front Genet 4: 289, 2013.
6. Goswami CP and Nakshatri H: PROGgene: Gene expression based survival analysis web application for multiple cancers. J Clin Bioinforma 3: 22, 2013.
7. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, et al: A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347: 1999-2009, 2002.
8. Pawitan Y, Bjöhle J, Amler L, Borg AL, Egyhazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, et al: Gene expression profiling spares early breast cancer patients from adjuvant therapy: Derived and validated in two population-based cohorts. Breast Cancer Res 7: R953-R964, 2005.
9. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 351: 2817-2826, 2004.
10. Wang J, Webb-Robertson BJ, Matzke MM, Varnum SM, Brown JN, Riensche RM, Adkins JN, Jacobs JM, Hoidal JR, Scholand MB, et al: A semiautomated framework for integrating expert knowledge into disease marker identification. Dis Markers 35: 513-523, 2013.
11. Lee G, Singanamalli A, Wang H, Feldman MD, Master SR, Shih NN, Spangler E, Rebbeck T, Tomaszewski JE and Madabhushi A: Supervised multi-view canonical correlation analysis (sMVCCA): Integrating histologic and proteomic features for predicting recurrent prostate cancer. IEEE Trans Med Imaging 34: 284-297, 2015.
12. Davis S and Meltzer PS: GEOquery: A bridge between the gene expression omnibus (GEO) and BioConductor. Bioinformatics 23: 1846-1847, 2007.
13. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, et al: Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365: 671-679, 2005.
14. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET and Bergh J: An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects and patient survival. Proc Natl Acad Sci USA 102: 13550-13555, 2005.
15. Ivshina AV, George J, Senko O, Mow B, Putti TC, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, et al: Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 66: 10292-10301, 2006.
16. Chanrion M, Negre V, Fontaine H, Salvetat N, Bibeau F, Mac Grogan G, Mauriac L, Katsaros D, Molina F, Theillet C and Darbon JM: A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. Clin Cancer Res 14: 1744-1752, 2008.
17. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al: Human protein reference database-2009 update. Nucleic Acids Res 37 (Database issue): D767-D772, 2009.
18. Nepusz T, Yu H and Paccanaro A: Detecting overlapping protein complexes in protein-protein interaction networks. Nat Methods 9: 471-472, 2012.
19. Friston KJ, Frith CD, Liddle PF and Frackowiak RS: Functional connectivity: The principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13: 5-14, 1993.
20. Gene Ontology Consortium: Gene Ontology Consortium: Going forward. Nucleic Acids Res 43 (Database issue): D1049-D1056, 2015.
21. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M and Tanabe M: Data, information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Res 42 (Database issue): D199-D205, 2014.
22. Parmar MB, Aliabadi HM, Mahdipoor P, Kucharski C, Maranchuk R, Hugh JC and Uludağ H: Targeting cell cycle proteins in breast cancer cells with siRNA by using lipid-substituted polyethylenimines. Front Bioeng Biotechnol 3: 14, 2015.
23. Weinberg RA: The retinoblastoma protein and cell cycle control. Cell 81: 323-330, 1995.
24. Yu Z, Baserga R, Chen L, Wang C, Lisanti MP and Pestell RG: microRNA, cell cycle, and human breast cancer. Am J Pathol 176: 1058-1064, 2010.
25. Tenga MJ and Lazar IM: Proteomic snapshot of breast cancer cell cycle: G1/S transition point. Proteomics 13: 48-60, 2013.
26. Achari C, Winslow S, Ceder Y and Larsson C: Expression of miR-34c induces G2/M cell cycle arrest in breast cancer cells. BMC Cancer 14: 538, 2014.
27. Kato T, Kameoka S, Kimura T, Tanaka S, Nishikawa T and Kobayashi M: p53, mitosis, apoptosis and necrosis as prognostic indicators of long-term survival in breast cancer. Anticancer Res 22: 1105-1112, 2002.
28. van Diest PJ, van der Wall E and Baak JP: Prognostic value of proliferation in invasive breast cancer: A review. J Clin Pathol 57: 675-681, 2004.
29. Königsberg R, Rögelsperger O, Jäger W, Thalhammer T, Klimpfinger M, De Santis M, Hudec M and Dittrich C: Cell cycle dysregulation influences survival in high risk breast cancer patients. Cancer Invest 26: 734-740, 2008.
30. Louie MC, McClellan A, Siewit C and Kawabata L: Estrogen receptor regulates E2F1 expression to mediate tamoxifen resistance. Mol Cancer Res 8: 343-352, 2010.
31. Kamel A, Mokhtar N, Elshakankiry N, Yassin D, Elnahass Y, Zakarya O, Elbasmy A and Elmetenawy W: The prognostic impact of some cell cycle regulatory proteins in Egyptian breast cancer patients. J Egypt Natl Canc Inst 18: 93-102, 2006.
32. Simon NE and Schwacha A: The Mcm2-7 replicative helicase: A promising chemotherapeutic target. Biomed Res Int 2014: 549719, 2014.
33. Coster G, Frigola J, Beuron F, Morris EP and Diffley JF: Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. Mol Cell 55: 666-677, 2014.
34. Labib K, Tercero JA and Diffley JF: Uninterrupted MCM2-7 function required for DNA replication fork progression. Science 288: 1643-1647, 2000.
35. Wan G, Hu X, Liu Y, Han C, Sood AK, Calin GA, Zhang X and Lu X: A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. EMBO J 32: 2833-2847, 2013.
36. de Munnik SA, Bicknell LS, Aftimos S, Al-Aama JY, van Bever Y, Bober MB, Clayton-Smith J, Edrees AY, Feingold M, Fryer A, et al: Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. Eur J Hum Genet 20: 598-606, 2012.
37. MacDermed DM, Khodarev NN, Pitroda SP, Edwards DC, Pelizzari CA, Huang L, Kufe DW and Weichselbaum RR: MUC1-associated proliferation signature predicts outcomes in lung adenocarcinoma patients. BMC Med Genomics 3: 16, 2010.
38. Wu G and Stein L: A network module-based method for identifying cancer prognostic signatures. Genome Biol 13: R112, 2012.
39. Cronin M, Sangli C, Liu ML, Pho M, Dutta D, Nguyen A, Jeong J, Wu J, Langone KC and Watson D: Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. Clin Chem 53: 1084-1091, 2007.