

## Original Article

# Transcription factor SLA2 regulated genes predict the survival of breast cancer patients

Liting Jin<sup>1</sup>, Yiping Gong<sup>2</sup>, Fangfang Chen<sup>1</sup>, Hongmei Zheng<sup>2</sup>, Manxiu Li<sup>2</sup>, Bin Xiong<sup>1</sup>

<sup>1</sup>Division of Surgical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, P. R. China; <sup>2</sup>Department of Breast Surgery, Hubei Cancer Hospital, Wuhan 430079, P. R. China

Received October 24, 2016; Accepted February 20, 2017; Epub March 1, 2017; Published March 15, 2017

**Abstract:** Objective: This study is to identify the molecules associated with the survival of breast cancer patients. Method: Gene expression and clinical information was obtained from GSE62944 with 1119 breast cancer samples and 113 normal samples. Cox scores were calculated with supervised principal components and the correlation between gene expression and patient survival was studied. The top 25 genes with max Cox scores were selected for patient classification and the patients were classified into 4 subgroups by ConsensusCluster. And 12019 differentially expressed genes and co-expression analysis was performed to found modules changed among the subgroups. Transcription factor (TF) regulation network was constructed so as to discover the key TFs associated with cancer survival. Results: The samples were divided 4 subgroups by 25 genes with the max Cox scores and survival curve of the 4 subgroups was significantly different, however, functional enrichment analysis showed no meaningful results for the 25 genes. Subgroups 1, 2, and 4 were chosen to screen molecular mechanisms underlying the different survival curve. Then a co-expression module, which contained genes that involved in adaptive immune response, was found to associate with the survival difference among the three subgroups. In addition, the genes in the module were regulated by TF SLA2 and could predict the outcome of breast cancer more robustness and sensitivity. Conclusion: Genes involved in adaptive immune response regulated by TF SLA2 are associated with the survival of breast cancer patients.

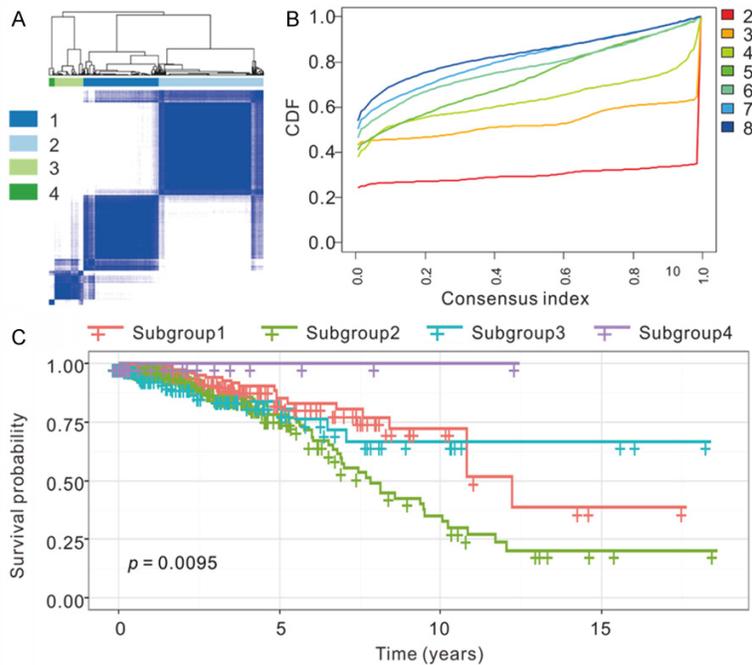
**Keywords:** Co-expression network, transcription factor regulation network, biomarker, breast cancer, differentially expressed genes

## Introduction

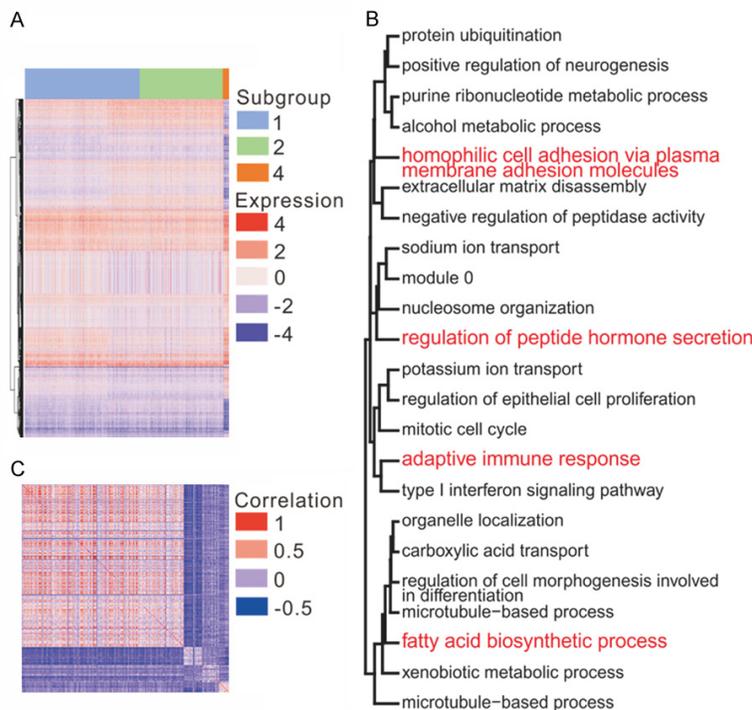
Breast cancer is a malignant tumor occurring in the epithelial tissue of mammary gland with 99% patients in women and only 1% in men. In the United States, about 234,190 new cases of breast cancer are estimated to be diagnosed and 40,730 Americans are expected to die from breast cancer in 2015 [1]. The 5-year relative survival rate for breast cancer is 89% in United States [1], however, in worldwide, the 5-year survival rate varies widely, from more than 80% in developed countries, such as Europe, America, and Japan, to about 60% in middle-income countries, and below 40% in low-income countries. To date, some factors have been proposed to increase breast cancer risk, such as obese, physical inactivity, alcohol consumption, etc [2, 3]. However, key molecular signaling pathways and factors determining breast cancer survival remain to be studied.

Given the considerable public health importance of breast cancer, various strategies are used for identifying key pathways related to patient outcome. The most commonly used method is to discover biomarkers, and single nucleotide polymorphism (SNP), gene/protein expression, lncRNA, miRNA, and methylation information have been mined for breast cancer biomarkers [4, 5]. A systematic review to investigate biomarkers obtained from 42 papers has been performed and find that biomarkers enrich in the response to steroid hormone stimulus and the cell cycle [6]. And an effective intervention in improving cancer outcomes and survival is used to monitor the crucial roles of insulin-like growth factors in breast cancer survivors [7]. An extensive analysis of signaling pathways find that the molecule p-p70S6K in mTOR pathway might be useful to predict clinical outcome of breast cancer patients [8].

## SLA2 predicts breast cancer patient survival



**Figure 1.** Tumor samples were classified into 4 subgroups with different survival curve significantly. A: The consensus matrices for  $k = 4$  obtained through PAM consensus clustering. B: The CDF curve for  $k = 2$  to 10. C: The survival curve of 4 subgroups with a  $P$  value 0.0095.

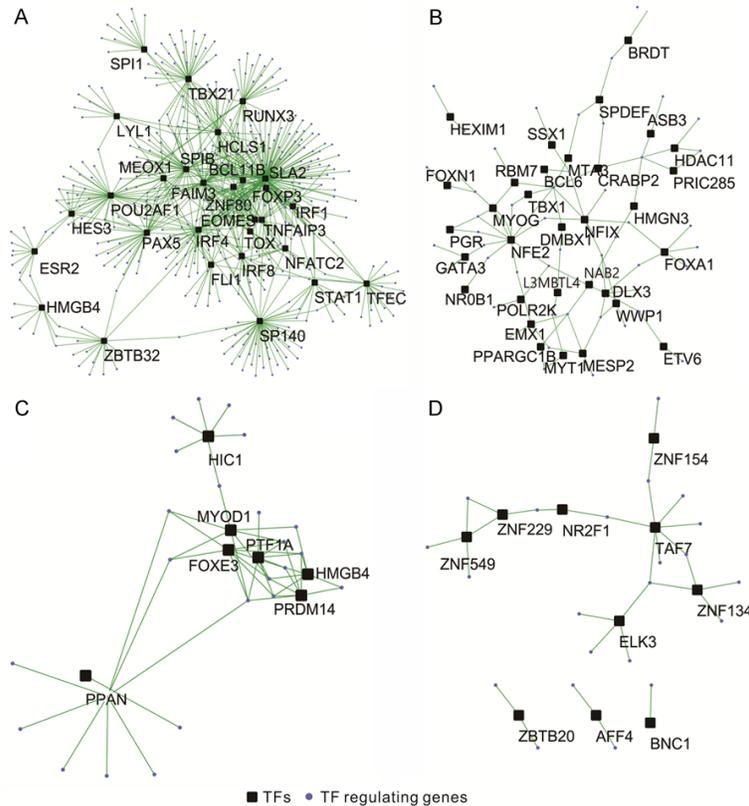


**Figure 2.** Co-expression analysis in DEGs found 4 significantly changed modules. A: DEGs between the 3 subgroups. B: The 23 co-expression modules and 4 significantly changed modules with red font. C: The correlation matrix of the genes involved in the 4 significantly changed modules.

In this study, transcriptome data of breast samples including 1232 samples were downloaded from public data. The significantly differentially expressed genes (DEGs), co-expression modules, and transcription factor regulation network between the high-risk patients with poor survival and the low-risk patients with good survival were integrated so as to unravel the mechanisms that are important in breast cancer survival.

### Materials and methods

The transcriptome data of breast samples from GSE-62944 including 1232 samples of 1119 tumors and 113 normal samples were downloaded and extracted [9]. Bioinformatic analysis was undertaken and all processes were performed using the R programming language (<https://www.r-project.org/>) together with its packages. The z-scores of all genes in the 1119 tumors were calculated with the following formula:  $z\text{-score}(X, Y) = [(value\ gene\ X\ in\ tumor\ Y) - (mean\ gene\ X\ in\ normal)] / (standard\ deviation\ X\ in\ normal)$ , in which X means a gene name and Y means one tumor sample. Then the R package superpc was used to calculate the Cox scores and sample classification was undertaken by Consensus Clusterplus package [10, 11]. DEGs were analyzed with limma package using the raw count [12]. Co-expression modules was performed by WGCNA and CVA package (<http://bioconductor.org/packages/devel/bioc/html/CVE.html>) [13]. Transcription factor regulation network was constructed with the RTN package [14]. Gene



**Figure 3.** The transcription factor regulation network of the 4 significantly changed modules. A: Adaptive immune response molecule. B: Fatty acid biosynthetic process molecule. C: Peptide hormone secretion regulation molecule. D: Homophilic cell plasma membrane adhesion molecule.

function enrichment was performed by clusterProfiler package [15]. Online tool (www.string-db.org) was applied for protein-protein interaction (PPI) network analysis [16], while external validation for new proposed biomarkers was processed by online tool SurvExpress [17].

**Results**

*Breast cancer was classified into 4 subgroups according to survival related genes*

To identify genes that related to survival in breast cancer, transcriptome data of 1232 samples, including 1119 tumors and 113 normal samples, from TCGA were collected and the z-scores for tumor samples were calculated. The Cox score for each gene in the datasets was measured and was then used to evaluate the correlation between gene expression and patient survival. The top 25 genes with max Cox scores were selected for patient classification and the patients were classified into 4 subgroups (Figure 1A and 1B). The status of

ER, PR and Her2 were presented in Supplemental File. The survive curve of the 4 subgroups was significantly different, however, functional enrichment analysis showed no meaningful results for the 25 genes and they did not provide hints for the mechanisms underlying the survival difference. This indicated that the biomarkers, which solely rely on their statistical associations with patient survival, might not provide enough biological information to illustrate the mechanisms of the survival of breast cancer patient. So the subgroups 1, 2, and 4 were selected for deep analysis to screen meaningful molecular features (Figure 1C) and explore the biological mechanisms explaining the outcome difference.

*Co-expression analysis found modules associated with different outcome*

In order to find co-expression modules that associate with patient survival, DEGs and co-expression analysis was undertaken. DEGs were firstly analyzed between the selected subgroups 1, 2, and 4 and a total of 12019 DEGs were found (Figure 2A). After the co-expression network was constructed, 23 co-expression modules were found in the DEGs. Four of these modules, including homophilic cell plasma membrane adhesion module, peptide hormone secretion regulation module, immune response adaptive module and fatty acid biosynthetic process module, were found to be different in the subgroups (Figure 2B). There are 560 genes involved in adaptive immune response, 36 genes belonging to homophilic cell plasma membrane adhesion molecule, 63 genes responsible for fatty acid biosynthetic process, and 59 genes regulating peptide hormone secretion (Figure 2C). To sum up, the results argued that integrated analysis of DEGs and co-expression network could narrow the search space for discovering meaningful modules.

## SLA2 predicts breast cancer patient survival

**Table 1.** The CI values of TF regulating genes derived from the online SurvExpress tools

TF*	Miller bergh breast GSE349-GPL96	Breast cancer meta-base: 10 cohorts 22k genes	BRCA-TCGA breast invasive carcinoma-July 2016	Kao Huang breast GSE20685	Breast invasive carcinoma TCGA
SLA2	0.96	0.84	0.85	0.93	0.95
FAIM3	0.76	0.70	0.74	0.75	0.75
IRF1	0.71	0.61	0.67	0.75	0.80
SP140	0.72	0.64	0.63	0.71	0.74
SPIB	0.70	0.66	0.67	0.80	0.73
NFE2	0.52	0.50	0.56	0.53	0.51
PPAN	0.59	0.54	0.58	0.60	0.61
TAF7	0.56	0.56	0.59	0.55	0.61
Top®	0.63	0.61	0.64	0.65	0.61

\*The TF regulating genes were used for SurvExpress analysis. ®Top are the 25 gens with max Cox scores for classification.

### TF network construction for the four modules

To find TFs that participate in gene regulation in the 4 modules and determination of breast cancer survival, TF network based on the whole dataset was constructed and genes in the aforementioned four modules had become a focus of concern. As for analysis of genes in the adaptive immune response co-expression module, there were 29 TFs with the minimum of 10 target regulating genes in the co-regulating pathway (**Figure 3A**). Two TFs, namely Src like adaptor 2 (SLA2) and interferon-regulatory factor (IRF), could regulate most genes independently. For the fatty acid biosynthetic process module, 32 TFs co-regulate the module together (**Figure 3B**). However, 26 of the 32 TFs could control 2-3 target genes in the module. By enrichment analysis, 6 TFs co-regulating genes in the peptide hormone secretion regulation module and 9 TFs in the hemophilic cell plasma membrane adhesion molecule were found (**Figure 3C** and **3D**). In total, the results demonstrated that TFs could regulate the co-expression modules and might play important roles in patient survival.

### Genes regulated by SLA2 could predict the outcome of breast cancer

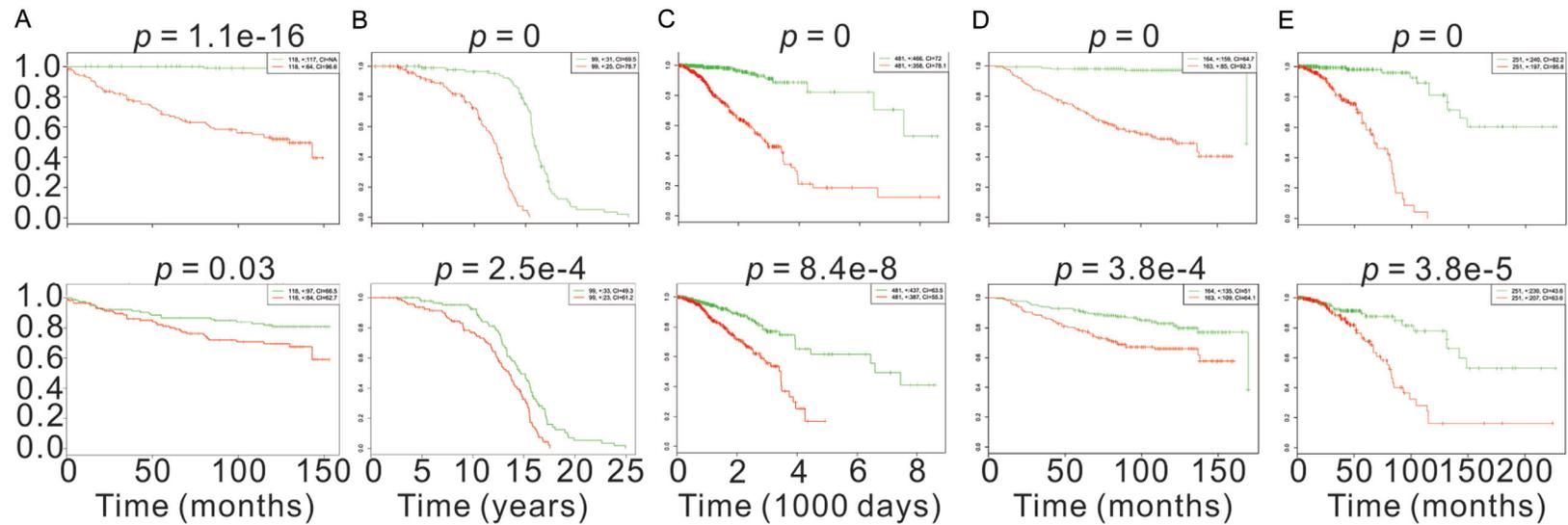
As mentioned above, TFs regulated genes in the four co-expression modules, but whether the TFs are associated with patient survival is not clear. In order to screen TFs regulated genes which could distinguish the outcome of breast patients, analysis of multiple breast cancer datasets was undertaken with SurvExpress online tools [17]. CI values and Log-rank

in the above found TFs were compared in the 5 datasets and 140 genes regulated by SLA2 were found to have better predictive results (**Table 1**). For SLA2-regulating genes, SurvExpress showed that the CI value was 0.91 (a mean of 5 datasets) (**Table 1**). We further compared the survival curve of the SLA2-regulating genes and the top 25 genes selected for classification in the above. The results also showed better predictive effect of SLA2-regulating genes, while the CI value of the top 25 genes was 0.63 (range from 0.61 to 0.65 in the 5 same datasets) (**Figure 4**). These findings suggested that SLA2 and its target regulating genes played important roles in the survival of breast cancer patients.

### SLA2 and its target regulating T-cell development

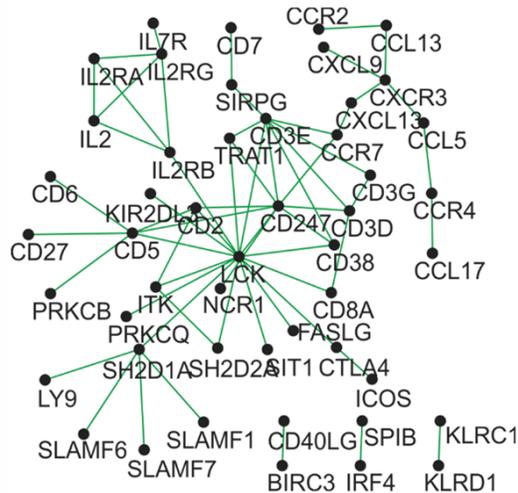
There were 140 genes predicted to be regulated by TF SLA2. To elaborate the biological functions associated with patient survival, Protein-Protein Interaction (PPI) network of the 140 genes together with SLA2 was obtained from the string-db. The interactions with evidence of experiments were shown in **Figure 5**. It was found that LCK, a member of the Src family of protein tyrosine kinases (PTKs), connected to other 16 proteins, and CD247 could bind to other 9 proteins. LCK protein is a key signaling molecule in the selection and maturation of developing T-Cells, and T-Cell receptor (TCR) and CD3 complex has been reported to mediate antigen recognition, ultimately resulting in T-Cell activation [18, 19]. SLA2 is a member of the SLAP family of adapter proteins,

## SLA2 predicts breast cancer patient survival



**Figure 4.** Comparing with the top 25 genes in max Cox scores, SLA2 regulating genes were better predictors of patient survival in 5 breast datasets. (A) The dataset of Miller Bergh Breast GSE349-GPL96. (B) The dataset of Breast Cancer Meta-base:10 cohorts 22K genes. (C) The dataset of BRCA-TCGA Breast invasive carcinoma-July 2016. (D) The dataset of Kao Huang Breast GSE20685 and (E) Breast Invasive Carcinoma TCGA.

## SLA2 predicts breast cancer patient survival



**Figure 5.** The protein-protein interaction (PPI) network analysis for SLA2 regulating genes. PPI network of SLA2 regulating genes was obtained from string-db. Interactions with no experimental evidence and genes with no interaction were not shown.

which plays an important receptor-proximal role in down-regulating T and B cell-mediated responses and inhibits antigen receptor-induced calcium mobilization [20, 21]. In conclusion, the results suggested that T-Cells played important roles in survival of breast cancer and SLA2 might be involved in the development of T-Cells.

### Discussion

Currently, most attentions are paid on DEGs analysis in carcinogenesis studies and DEGs are generally selected to mine signaling pathway at first and the pathways identified are then used to reveal molecular features under pathology [22]. However, DEGs are the consequences induced by specific pathological phenomena other than the causes of phenotypic changes. It has been well accepted that cancer is caused by genomic changes, such as somatic mutations, copy number variation, structural variation, and methylation changes, which then resulting in a large number of DEGs and abnormal signaling pathways [23]. That means that DEGs are induced by a certain diseases or phenotypes, but not the causal mechanisms. Therefore, only analysis of DEGs and the subsequent functional enrichment have no chance to uncover key regulatory factors. In contrast, co-expression analysis and transcription factor regulation network inference, which are useful

supplements to the analysis of DEGs in transcriptome datasets in recent years, have been used to discover the regulated mechanisms underlying the outcome [24-26].

In the present study, using the method of integrating DEGs, gene co-expression analysis, and transcription regulating network inference, we identified that the transcription factor SLA2-regulating genes, which enriched in adaptive immune response and were involved in T-Cell development and activation, were associated with the survival of breast cancer patients. The findings confirmed the important roles of T-Cell in cancer survival and proposed the potential target for drug design.

At the beginning, we identified 25 genes according to correlations between their expressions with outcome, and then classified the patients into 4 subgroups with different survival curve. However, the 25 genes were identified solely relying on their statistical associations with patient survival, which, inefficiently applied the co-expression and regulation information of the high throughput data. As a result, these genes had no meaningful biological information related to pathogenic mechanisms.

In the second step of this study, DEGs, gene co-expression analysis, and transcription regulating network inference were integrated to find the key regulators determining patient survival. DEG analysis found 12019 genes and co-expression analysis found 23 modules, and 4 modules were different significantly between the subgroups. The genes in the differential co-expression modules represented the difference and correlation between the samples simultaneously and provided more information of the dynamic changes in the gene regulatory networks. In addition, these genes may be coordinated in order to perform related cellular functions and correlated with specific conditions. This provides an opportunity to identify key regulators, such as TFs, to illuminate the mechanisms of specific disease phenotype. We constructed the transcription factor regulation network and found the important roles of TF SLA2 in regulating the differential module of adaptive immune response.

In conclusion, the combination of DEGs, co-expression network, and transcription factor regulation network inference was able to pre-

dict only one transcription factor determining the survival of breast cancer patients. Our research investigated the molecular signatures for breast cancer at transcriptional regulation level and confirmed the roles of T-cell development in outcome determining. The integrated method can also be applied to other cancer researches for the discovering of the molecular mechanisms behind survival curve.

**Disclosure of conflict of interest**

None.

**Address correspondence to:** Bin Xiong, Division of Surgical Oncology, Zhongnan Hospital of Wuhan University, No. 169, Donghu Road, Wuhan 430071, P. R. China. Tel: +86-18607115962; E-mail: Binxiong1961@gmail.com

**References**

[1] American Cancer Society. Cancer Facts and Figures. 2015.

[2] Colditz GA. From epidemiology to cancer prevention: implications for the 21st Century. *Cancer Causes Control* 2007; 18: 117-123.

[3] Chlebowski RT, Manson JE, Anderson GL, Cauley JA, Aragaki AK, Stefanick ML, Lane DS, Johnson KC, Wactawski-Wende J, Chen C, Qi L, Yasmeen S, Newcomb PA and Prentice RL. Estrogen plus progestin and breast cancer incidence and mortality in the women’s health initiative observational study. *J Natl Cancer Inst* 2013; 105: 526-535.

[4] Zhang K, Luo Z, Zhang Y, Zhang L, Wu L, Liu L, Yang J, Song X and Liu J. Circulating lncRNA H19 in plasma as a novel biomarker for breast cancer. *Cancer Biomark* 2016; 17: 187-194.

[5] Zavala V, Perez-Moreno E, Tapia T, Camus M and Carvallo P. miR-146a and miR-638 in BRCA1-deficient triple negative breast cancer tumors, as potential biomarkers for improved overall survival. *Cancer Biomark* 2016; 16: 99-107.

[6] Abba MC, Lacunza E, Butti M and Aldaz CM. Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. *Biomark Insights* 2010; 5: 103-118.

[7] Meneses-Echavez JF, Jimenez EG, Rio-Valle JS, Correa-Bautista JE, Izquierdo M and Ramirez-Velez R. The insulin-like growth factor system is modulated by exercise in breast cancer survivors: a systematic review and meta-analysis. *BMC Cancer* 2016; 16: 682.

[8] Horii R, Matsuura M, Dan S, Ushijima M, Uehiro N, Ogiya A, Honma N, Ito Y, Iwase T, Yamori T

and Akiyama F. Extensive analysis of signaling pathway molecules in breast cancer: association with clinicopathological characteristics. *Int J Clin Oncol* 2015; 20: 490-498.

[9] Rahman M, Jackson LK, Johnson WE, Li DY, Bild AH and Piccolo SR. Alternative preprocessing of RNA-sequencing data in The Cancer Genome Atlas leads to improved analysis results. *Bioinformatics* 2015; 31: 3666-3672.

[10] Bair E and Tibshirani R. Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol* 2004; 2: E108.

[11] Wilkerson MD and Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; 26: 1572-1573.

[12] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; 43: e47.

[13] Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.

[14] Castro MA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, Tilley WD, Markowitz F, Ponder BA and Meyer KB. Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat Genet* 2016; 48: 12-21.

[15] Yu G, Wang LG, Han Y and He QY. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012; 16: 284-287.

[16] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ and von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; 43: D447-452.

[17] Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, Martinez-Torteya A, Chacolla-Huaringa R, Rodriguez-Barrientos A, Tamez-Pena JG and Trevino V. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. *PLoS One* 2013; 8: e74250.

[18] Isakov N and Biesinger B. Lck protein tyrosine kinase is a key regulator of T-cell activation and a target for signal intervention by Herpesvirus saimiri and other viral gene products. *Eur J Biochem* 2000; 267: 3413-3421.

[19] Pigeon SV, Tabarin T, Yamamoto Y, Ma Y, Bridgeman JS, Cohnen A, Benzing C, Gao Y, Crowther MD, Tungatt K, Dolton G, Sewell AK, Price DA, Acuto O, Parton RG, Gooding JJ, Rossy J, Rossjohn J and Gaus K. Functional role of T-cell receptor nanoclusters in signal ini-

## SLA2 predicts breast cancer patient survival

- tiation and antigen discrimination. *Proc Natl Acad Sci U S A* 2016; 113: E5454-5463.
- [20] Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, Kim H, Kim K, Yang S, Bae D, Yun A, Kim S, Kim CY, Cho HJ, Kang B, Shin S and Lee I. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep* 2015; 5: 11432.
- [21] Holland SJ, Liao XC, Mendenhall MK, Zhou X, Pardo J, Chu P, Spencer C, Fu A, Sheng N, Yu P, Pali E, Nagin A, Shen M, Yu S, Chan E, Wu X, Li C, Woisetschlager M, Aversa G, Kolbinger F, Bennett MK, Molineaux S, Luo Y, Payan DG, Mancebo HS and Wu J. Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling. *J Exp Med* 2001; 194: 1263-1276.
- [22] Khatri P, Sirota M and Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 2012; 8: e1002375.
- [23] Stratton MR, Campbell PJ and Futreal PA. The cancer genome. *Nature* 2009; 458: 719-724.
- [24] de la Fuente A. From 'differential expression' to 'differential networking'-identification of dysfunctional regulatory networks in diseases. *Trends Genet* 2010; 26: 326-333.
- [25] Yu H, Liu BH, Ye ZQ, Li C, Li YX and Li YY. Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC Bioinformatics* 2011; 12: 315.
- [26] Udyavar AR, Hoeksema MD, Clark JE, Zou Y, Tang Z, Li Z, Li M, Chen H, Statnikov A, Shyr Y, Liebler DC, Field J, Eisenberg R, Estrada L, Massion PP and Quaranta V. Co-expression network analysis identifies Spleen Tyrosine Kinase (SYK) as a candidate oncogenic driver in a subset of small-cell lung cancer. *BMC Syst Biol* 2013; 7 Suppl 5: S1.