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

Liting Jin¹, Yiping Gong², Fangfang Chen², Hongmei Zheng², Manxiu Li², Bin Xiong¹

¹Division of Surgical Oncology, Zhongnan Hospital of Wuhan University, Wuhan 430071, P. R. China; ²Department 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].
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: 
\[ z-score (X, Y) = \frac{(value \text{ gene } X \text{ in tumor } Y - mean \text{ gene } X \text{ in normal})}{standard \text{ deviation } X \text{ in normal}} \], 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 were 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 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.
SLA2 predicts breast cancer patient survival

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

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

*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

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.

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.

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

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.

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

SLA2 predicts breast cancer patient survival


