Application of the Shortest Path Algorithm for the Discovery of Breast Cancer-Related Genes

Lei Chen, ZhiHao Xing, Tao Huang, Yang Shu, GuoHua Huang and Hai-Peng Li

1College of Information Engineering, Shanghai Maritime University, Shanghai 201306, People’s Republic of China
2The Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Jiaotong University School of Medicine and Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200025, People’s Republic of China
3Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4State Key Laboratory of Medical Genomics, Institute of Health Sciences, Shanghai Jiaotong University School of Medicine and the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200025, People’s Republic of China
5Institute of Systems Biology, Shanghai University, Shanghai 200444, People’s Republic of China
6CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, People’s Republic of China

Abstract: Breast cancer, the most prevalent cancer in women, develops from breast tissue. Its incidence has increased in recent years due to environmental risk factors. Thus, it is urgent to uncover the mechanism underlying breast cancer to design effective treatments. Identification of all breast cancer-related genes is one way to help elucidate the underlying breast cancer mechanism. In this study, a computational method was built and applied to discover new candidate breast cancer-related genes. Based on the known breast cancer-related genes retrieved from public databases, the shortest path algorithm was applied to discover new candidate genes in the protein-protein interaction network. The analysis results of the selected genes suggest that some of them are deemed breast cancer-related genes according to the most recent published literature, while others have direct or indirect associations with the initiation and development of breast cancer.

Keywords: Betweenness, breast cancer, disease gene, protein-protein interaction, shortest path algorithm, weighted network.

1. INTRODUCTION

Breast cancer is a malignant tumor that develops from breast tissue. It contains a cluster of cancer cells that not only can invade surrounding tissues but also spread to other parts of the body, such as the regional lymph nodes, lungs, liver, and bone-marrow [1]. Most cases of breast cancer occur in women; however, rarely, men could also acquire the cancer, which only accounts for a small percentage of all male tumors. In recent years, the incidence of this disease has increased [2]. Breast cancer, which is the most prevalent cancer in women, accounts for 22.9% of invasive cancers and 16% of all cancers in females. In 2008, 458,503 deaths were caused by breast cancer worldwide, resulting in 13.7% cancer deaths in women [3]. The incidence of breast cancer in developed nations is higher than that in developing countries. Possible contributory factors include lifestyle and eating habits [4].

Breast cancer is a complicated disease with different biological features. Although it is difficult to determine whether an individual could develop breast cancer, several hereditary and environmental risk factors have been reported to affect the likelihood of developing the disease; among them, female gender and older age [5] are the primary risk factors. Other factors, such as lifestyle, obesity, childbearing [6], estrogen exposure [7], radiation exposure [8] and genetic factors, have also reported to have associations with the formation and development of breast cancer.

It is believed that breast cancer results from cumulative genetic damage and genetic alterations, resulting in the activation of proto-oncogenes and inactivation of tumor suppressor genes. Genetic factors have been observed in one-fourth of breast cancer patients [9]. Approximately 5-10% of breast cancer cases may be due to mutations in high susceptibility genes, such as \textit{BRCA1} and \textit{BRCA2} [10]. \textit{BRCA1} and \textit{BRCA2} genes encode human tumor suppressor proteins that play roles in repairing damaged DNA and ensure genome stability. When mutations occur in these genes, BRCA protein cannot function properly. Mutations in \textit{BRCA1} and \textit{BRCA2} increase the risk of female breast cancer and account for approximately 5 to 10 percent of breast cancers [11]. There are several other genes that are also
associated with breast cancer, such as PTEN, TP53 and ATM. The PTEN gene is involved in the regulation of cell growth, apoptosis and metastasis [12,13]. Defective PTEN protein causes cells to divide in an uncontrolled way and can render people to a higher risk of cancerous breast tumors. The TP53 gene encodes p53 protein. The p53 protein is crucial in humans because it regulates the cell cycle and functions as a tumor suppressor. People with this rare syndrome have a higher risk of breast cancer and several other cancers [14]. The ATM gene can also help repair damaged DNA and phosphorylate several key proteins that initiate the activation of the DNA damage checkpoint in DNA repair or apoptosis [15]. Patients with ATM mutations have an increased risk for breast cancers [16].

Known as one of the most common cancers, the concrete mechanism underlying breast cancer has not yet been completely understood. Because breast cancer is a complex disease, and the number of human genes is huge, it is difficult to discover novel breast cancer-related genes by experiments only. By contrast, bioinformatics methods, which have been applied to tackle various disease-related problems [17-23], can provide an alternative way to efficiently screen quantities of breast cancer-related genes at the same time. Using certain bioinformatics methods [24, 25], the associations between the selected genes and breast cancer can be analyzed, thereby measuring the likelihood of them being novel breast cancer-related genes.

This study applied an existing bioinformatics method, which has been used to investigate related genes of age-related macular degeneration [18] and hepatocellular carcinoma [23], to identify novel breast cancer-related genes based on known ones. According to certain studies [26-30], proteins that can interact with each other always share similar functions. Thus, a weighted network was constructed based on the information of protein-protein interactions retrieved from STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) [31]. By application of the shortest path algorithm to this network to identify all of the shortest paths connecting any two known breast cancer-related genes, genes occurring in at least one shortest path were selected as candidate genes. Next, a randomization test was used to filter these candidate genes. To analyze the biological functions of the final selected genes, Gene Ontology (GO) and the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway were employed that have been widely used to annotate enrichment and analyze the functions of genes [32, 33]. The GO and KEGG pathway enrichment analysis of the candidate genes indicates that certain GO terms and KEGG pathways are highly associated with the development of breast cancer. Further analysis suggests that some candidate genes have been reported as breast cancer-related genes in some of the most recently published literature. We hope that our results would help to uncover the mechanism of breast cancer and provide new insights into novel therapies.

2. MATERIALS AND METHODS

2.1. Materials

The breast cancer-related genes were collected from the following two databases: UniProtKB (Protein Knowledgebase, http://www.uniprot.org/uniprot/, Release 2013_12) [34] and TSGene Database (Tumor Suppressor Gene Database, http://bioinfo.mc.vanderbilt.edu/TSGene/cancer_type.cgi) [35]. In detail, 224 reviewed genes and 130 tumor suppressor genes that are related to breast cancer were retrieved from UniProtKB, whereas 154 breast cancer-related genes were obtained from the TSGene Database. All of the obtained genes were combined. Finally, we obtained 369 breast cancer-related genes, which are available in the Online Supporting Information S1.

2.2. Methods to Identify Novel Genes and the Screen Process

The original idea of the method was based on the fact that proteins that can interact with each other always share similar functions [26-30]. Thus, the information of protein-protein interactions, including direct (physical) and indirect (functional) interactions, retrieved from STRING [31], was employed. These interactions are derived from genomics, high-throughput experiments, (conserved) coexpression and previous knowledge. In the obtained file, each of the obtained protein-protein interactions contains two proteins and one score. The score roughly estimates how likely a given interaction describes a functional linkage between two proteins. Obviously, the higher the score of the interaction is, the stronger the functional linkage is. For formulation, \( Q(p_1, p_2) \) is considered the score of the interaction between proteins \( p_1 \) and \( p_2 \).

The model for the discovery of novel breast cancer-related genes was built based on the information of protein-protein interactions. Similar models have been applied for the identification of related genes of other diseases, such as age-related macular degeneration [18] and hepatocellular carcinoma [23]. Here, we provided a brief description of the method, and readers can refer to the studies of Zhang et al. and Jiang et al. [18, 23] for detail.

1. According to the information of protein-protein interaction retrieved from STRING, a weighted network was constructed by taking proteins as nodes; two nodes were adjacent if and only if the corresponding proteins can interact with each other. Because the score of each interaction ranged between 150 and 999, each edge with end-nodes \( v_1 \) and \( v_2 \) was assigned a weight defined as 1000-\( Q(p_1, p_2) \), where \( p_1 \) and \( p_2 \) are corresponding proteins of nodes \( v_1 \) and \( v_2 \).

2. All of the shortest paths in the above constructed network, which connected any two known breast cancer-related genes, were obtained by the well-known shortest path algorithm, Dijkstra’s algorithm [36], which is integrated into Maple.

3. For each node/gene in the network, the number of shortest paths that contained the node/gene as an inner node was counted. This value was called betweenness in the present study. Genes with betweennesses greater than zero were picked up as candidate genes, which would be further considered in the following procedures.

4. To avoid false discoveries, we randomly selected 500 gene sets whose sizes were equal to the size of the
gene set consisting of known breast cancer-related genes. Next, we calculated the betweenness for each candidate gene from each of these 500 gene sets.

5. The permutation FDR was calculated for each candidate gene, which was defined as “the number of gene sets on which the betweennesses were larger than the betweenness of the known breast cancer-related gene set”/500. Genes with small permutation FDRs were picked up as significant candidate genes.

3. RESULTS AND DISCUSSION

3.1. Candidate Genes

According to the method, all of the shortest paths connecting any two known breast cancer-related genes were searched in the weighted network constructed by the information of protein-protein interactions. The betweenness of each node/gene in the network was computed based on these paths. The value of a certain gene’s betweenness indicates the strength of the direct and indirect associations between the gene and known breast cancer-related genes [37]. In detail, a high betweenness suggests a strong association, whereas a low betweenness suggests a weak association. Thus, we selected 621 genes whose betweennesses were greater than zero and termed these genes as candidate genes. These genes were listed in the Online Supporting Information S2.

Until now, we obtained 621 candidate genes for breast cancer. However, some of them may be false discoveries because they may have special positions in the protein-protein interaction network, resulting in high betweenness in each case. It is necessary to execute a randomization test mentioned in the fourth and fifth steps. As a result, we obtained the permutation FDR for each of the 621 candidate genes and listed these permutation FDRs in Online Supporting Information S2. To exclude false discoveries, we selected 62 genes with permutation FDRs smaller than 0.01 as significant candidate genes. These 62 genes and their betweennesses are listed in Table 1. In the following sections, we analyzed the likelihood of these genes to be novel breast cancer-related genes.

3.2. Results from DAVID

As mentioned in Section 3.1, 62 genes were obtained as significant candidate genes for breast cancer. To analyze the relationship between them and breast cancer, the functional annotation tool DAVID (Database for Annotation, Visualization and Integrated Discovery) [38] was employed to understand the biological meaning underlying these 62 genes. In fact, DAVID analyzed the enrichments of 62 genes on GO terms or KEGG pathways, thereby inferring the relationship between genes and some biological processes. This method has been applied to study various disease-related and gene-related problems [18, 23, 39, 40]. The analysis results in this study can be found in Online Supporting Information S3. Thus, the next two sections provided the detailed analysis of 62 genes on GO terms and KEGG pathways, respectively.

3.2.1. GO Term Analysis

From Online Supporting Information S3, 122 GO terms were found to be enriched by the 62 genes. We analyzed and discussed the top ten GO terms sorted by P-value, including seven biological process (BP) GO terms, two cellular component (CC) GO terms and one molecular function (MF) GO term, respectively. It is necessary to note that the P-values of these GO terms were all less than 0.05, indicating that the 62 significant candidate genes were highly enriched in these GO terms. Fig. (1) shows these ten GO terms and the ‘count’ obtained by DAVID that is defined as the number of genes among the 62 genes that shared these GO terms.

The seven BP terms are (I) GO: 0010604 (positive regulation of macromolecule metabolic process) (“count”=14); (II) GO: 0010557 (positive regulation of macromolecule biosynthetic process) (“count”=11); (III) GO: 0031328 (positive regulation of cellular biosynthetic process) (“count”=11); (IV) GO: 0009891 (positive regulation of biosynthetic process) (“count”=11); (V) GO: 0051276 (chromosome organization) (“count”=9); (VI) GO: 0016569 (covalent chromatin modification) (“count”=5); and (VII) GO: 0045941 (positive regulation of transcription) (“count”=9).

![Fig. (1). The top ten GO terms enriched by 62 genes sorted by P-value. The X-axis represents the GO term ID, while the Y-axis represents the number of genes among the 62 genes that shared the GO terms.](image-url)
The macromolecule biosynthetic process is relatively important in the development of breast cancer. Previous studies have shown that the biosynthesis of macromolecules, including carbohydrates, proteins, lipids and nucleic acids, needed to be altered to support rapid growth and survival of tumor cells [41]. Moreover, up-regulated macromolecule metabolism was found after treating breast cancer with chemotherapy, providing mechanistic insights into the resistance to chemotherapy and revealing the possible role of the macromolecule biosynthetic process in breast cancer [42]. The next two terms represented the importance of the biosynthetic process in breast cancer. Like the macromolecule biosynthesis, an increased biosynthetic process can provide rapid ATP generation and increased biosynthesis of macromolecules such as nucleic acids and hormones for tumor cells [41]. In cancer cell, biosynthetic pathways are altered toward an anabolic metabolism to meet the needs of cell proliferation [43]. All of the above indicated the relationship between biosynthetic pathways and breast cancer and may help us to explore new mechanisms in breast cancer research.

<table>
<thead>
<tr>
<th>Ensemble ID</th>
<th>Gene Name</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSP00000227507</td>
<td>CCND1</td>
<td>2697</td>
</tr>
<tr>
<td>ENSP00000250720</td>
<td>SMARC4</td>
<td>1523</td>
</tr>
<tr>
<td>ENSP00000262158</td>
<td>SMAD7</td>
<td>1185</td>
</tr>
<tr>
<td>ENSP00000220005</td>
<td>CDC37</td>
<td>1041</td>
</tr>
<tr>
<td>ENSP00000358918</td>
<td>SUFU</td>
<td>687</td>
</tr>
<tr>
<td>ENSP0000028916</td>
<td>SCNN1A</td>
<td>684</td>
</tr>
<tr>
<td>ENSP0000034392</td>
<td>XRCC3</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000364403</td>
<td>UBR4</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000370867</td>
<td>TGM3</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000317039</td>
<td>RMI1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000262741</td>
<td>PIK3R3</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000217026</td>
<td>MYBL2</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000337354</td>
<td>LIPA</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000322180</td>
<td>DSCC1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP0000039723</td>
<td>CIR1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000306522</td>
<td>CAMTA1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000352011</td>
<td>CACNA1G</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000357040</td>
<td>VANGL2</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000356591</td>
<td>SOAT1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000349959</td>
<td>RICTOR</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000355245</td>
<td>PAX9</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000353344</td>
<td>ETS2</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000280362</td>
<td>PTS</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000276927</td>
<td>IFNA1</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000346566</td>
<td>CKAP5</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000323076</td>
<td>NDUFAF3</td>
<td>343</td>
</tr>
<tr>
<td>ENSP00000347232</td>
<td>BLM</td>
<td>343</td>
</tr>
<tr>
<td>ENSP0000034854</td>
<td>ZACN</td>
<td>342</td>
</tr>
<tr>
<td>ENSP00000316948</td>
<td>CLK4</td>
<td>341</td>
</tr>
<tr>
<td>ENSP00000262643</td>
<td>CCNE1</td>
<td>115</td>
</tr>
<tr>
<td>ENSP00000345785</td>
<td>FZD9</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Detailed information of the 62 significant candidate genes.
these processes has been related to tumor initiation and development. The mechanisms of chromatin remodeling involve covalent histone modifications, DNA methylation and many other processes [44, 45]. For example, the relationship between the overexpression of EZH2, a methyltransferase for H3K27, and tumor progression, has been observed in breast cancer [46]. These indicate that chromatin plays a fundamental role and underlies many other processes in breast cancer.

The inclusion of GO: 0045941 (positive regulation of transcription) meets our expectations. As an important part of the central dogma, transcription is involved in almost all of the cellular processes, and the changes in expression of many genes have been observed in breast cancer [47, 48].

The two CC terms are (I) GO: 0005654 (nucleoplasm) (“count”=13); and (II) GO: 0005829 (cytosol) (“count”=14). These revealed the role of the cytosol and nucleoplasm in the development of breast cancer, providing some suggestions about how breast cancer occurs.

The only MF term is GO: 0019207 (kinase regulator activity) (“count”=5). The relationship between GO: 0019207 and breast cancer cell lines has been revealed in one previous study [49]. Additionally, kinase activity has played an important role in fundamental cellular processes, and disruption of kinase activity may be related to cancer. In fact, MAP kinase, as an important signal molecule, is highly associated with breast cancer growth and apoptosis [50].

3.2.2. KEGG Pathway Analysis

From Online Supporting Information S3, 13 KEGG pathways were found to be enriched by these 62 genes. Fig. (2) shows the 13 KEGG pathways and the ‘count’ obtained by DAVID that is defined as the number of genes among the 62 genes that shared these pathways.

From Fig. (2), these 13 KEGG pathways are (I) hsa05200 (Pathways in cancer) (“count”=8); (II) hsa05212 (Pancreatic cancer) (“count”=4); (III) hsa05220 (Chronic myeloid leukemia) (“count”=4); (IV) hsa04310 (Wnt signaling pathway) (“count”=5); (V) hsa05210 (Colorectal cancer) (“count”=4); (VI) hsa05222 (Small cell lung cancer) (“count”=4); (VII) hsa05219 (Bladder cancer) (“count”=3); (VIII) hsa05223 (Non-small cell lung cancer) (“count”=3); (IX) hsa05214 (Glioma) (“count”=3); (X) hsa04115 (p53 signaling pathway) (“count”=3); (XI) hsa05218 (Melanoma) (“count”=3); (XII) hsa05215 (Prostate cancer) (“count”=3); (XIII) hsa00100 (Steroid biosynthesis) (“count”=2).

These 62 genes are enriched in many obvious cancer KEGG pathways, including hsa05200 (Pathways in cancer), hsa05212 (Pancreatic cancer), hsa05220 (Chronic myeloid leukemia), hsa05210 (Colorectal cancer), hsa05222 (Small cell lung cancer), hsa05219 (Bladder cancer), hsa05223 (Non-small cell lung cancer), hsa04115 (p53 signaling pathway), hsa05214 (Glioma), hsa05218 (Melanoma), and hsa05215 (Prostate cancer), indicating that these genes are highly associated with tumor initiation and development and may play the same roles in breast cancer. Aberrant regulation of the Wnt signaling pathway has been found in breast cancer [51]. Blocking the Wnt signaling pathway with iCRT-3 inhibitor and SOX4 knockdown resulted in the inhibition of cell proliferation and induced apoptosis in TNBC (triple-negative breast cancer) [52], suggesting an internal relationship between the Wnt signaling pathway and breast cancer. Steroid biosynthesis also plays an important role in breast cancer. For example, prolonged exposure to estrogen, a steroid, increased the risk of breast cancer possibly by facilitating the proliferation of breast cells [53].

3.3. Analysis of Some Significant Candidate Genes

Table 1 lists the 62 significant candidate genes obtained by our method. Among them, 11 genes had a betweenness larger than 680, while the betweennesses of the remaining genes were less than 400. It is a great gap, indicating that 11 genes may be acute breast cancer-related genes with a higher possibility than others. In addition, four genes—CCND1, LRP6, TGFBR1 and SMAD7—have been recognized as breast cancer-related genes according to the most recent published literature, whereas the other 7 genes were highly

![Fig. (2). The 13 KEGG pathways enriched by 62 genes. The X-axis represents the KEGG pathway ID, while the Y-axis represents the number of genes among the 62 genes that shared the KEGG pathways.](image-url)
associated with the initiation and development of tumors and may provide evidence for further research of breast cancer. The following paragraphs provide the detailed analyses.

**CCND1.** The betweenness of this gene was 2,697, which was the highest among the 62 significant candidate genes. The **CCND1** gene encodes the protein cyclin-D1, which belongs to the highly conserved cyclin family. **CCND1** exhibits periodical expression across the cell cycle and can regulate cyclin-dependent kinase. Different cyclins work together to control the entire cell cycle. Mutations and amplification of the **CCND1** gene are observed frequently in many types of tumors and alter the normal cell cycle, possibly contributing to tumorigenesis [54, 55]. Moreover, overexpression of cyclin D1 has been found in breast cancer and may serve as a marker for metastasis in clinical treatment [56].

**LRP6.** **LRP6** had a betweenness of 1,868, which was the second highest value. The protein encoded by **LRP6** is a low-density lipoprotein (LDL) receptor. LDL receptors, which are located on the cell surface, play an important role in the endocytosis of lipoprotein. Through the interaction with the Wnt signal pathway, **LRP6** is involved in the regulation of proliferation and migration in cancer. Previous research has shown that Wnt signaling activation caused by overexpression of **LRP6** can contribute to the tumorigenesis of breast cancer [57].

**SMARCA4.** **SMARCA4**, with a betweenness of 1,523 (third highest), encodes a member of the SWI/SNF family. It is related to several important tumor suppressor proteins, and mutations are found in many cancer cell lines such as breast cancer [58].

**TGFBR1.** **TGFBR1** showed the fourth highest betweenness value (1,486). **TGFBR1** and type II TGF-beta receptors together form a heteromeric complex, which can transduce the TGF-beta signal from the membrane to cytoplasm. The **TGFBR1*6A** variant has been associated with a high risk for breast cancer, so a better understanding of the biological function of **TGFBR1** signaling may help to assess breast cancer risk and prevent breast cancer [59].

**SMAD7.** The betweenness of **SMAD7** was 1,185 (the fifth highest). In breast tumors, **SMAD7** can act as a negative regulator of TGF-β, and it is reported that the formation of bone metastases is inhibited by the overexpression of Smad7 [60, 61].

**CDK4.** The betweenness of this gene was 1,060, which was the sixth highest value. **CDK4** encodes a protein belonging to the Ser/Thr protein kinase family. **CDK4** was involved in a protein kinase complex that controls the G1/S phase transition [62]. A previous report has shown that the maintenance of breast cancer requires the presence of **CDK4** activity, and clinical therapies targeting **CDK4** kinase activity in cancer may be promising [63].

**CDC37.** The betweenness of **CDC37** was 1,041 (the seventh highest). **CDC37** is a molecular chaperone that can form a complex with Hsp90 and many protein kinases such as **CDK4**. Moreover, **CDC37**/Hsp90 was reported to contribute to the stabilization of newly synthesized **CDK4** [64].

**MMP9.** The betweenness of **MMP9** was 883, which was the eighth highest value. **MMP9** is a member of the matrix metalloproteinase (MMP) family that has been indicated in many biological processes, including reproduction, embryonic development and cancer metastasis. The role of **MMP** in breast cancer has been diverse. Unlike several MMPs that come from stromal cells, **MMP9** is mainly produced by cancer cells in breast cancer. Additionally, in a mouse breast cancer model, **MMP9** forms tumor cells that are shown to be required for invasion and pulmonary metastasis [65].

**SUFU.** **SUFU** received the ninth highest betweenness value (687). Suppressor of fused (**SUFU**) acts as a negative regulator of the Hedgehog signaling pathway by binding to Gli [66]. The primary function of the Hedgehog signaling pathway (Hh) is restraint in embryogenesis, except for several processes such as adult tissue repair, and aberrant reactivation of Hh has been associated with several types of cancers [67]. Thus, as a negative regulator, **SUFU** may play a role in breast cancer.

**SCNN1A.** The betweenness of this gene was 684, which was the tenth highest value. **SCNN1A** is involved in the formation of sodium channels, which control fluid and electrolyte transport across epithelia.

**SET.** The betweenness of this gene was 684, which was the eleventh highest value. The encoded protein **SET** is an ATP-dependent histone acetyltransferase that is involved in the regulation of several processes such as cell cycle progression, gene expression, and chromatin remodeling. The role of **SET** in breast cancer has been less understood, but it is believed that its overexpression may contribute to the development of breast cancer [68, 69].

**CONCLUSION**

The discovery of disease-related genes is an important research area in biomedicine and genomics. This study applied an existing computational method to discover new breast cancer-related genes. The results show that this method is effective for tackling this problem. We hope some of the newly discovered genes will be confirmed by solid experiments.

**CONFLICT OF INTEREST**

The authors confirm that this article content has no conflict of interest.

**ACKNOWLEDGEMENTS**

This work was supported by the National Science Foundation of China (61202021, 61373028, 11371008, 61303099), Innovation Program of Shanghai Municipal Education Commission (12YZ120), Shanghai Educational Development Foundation (12CG55).

**SUPPLEMENTARY MATERIAL**

Supplementary material is available on the publisher’s web site along with the published article.

**REFERENCES**

Algorithm for the Discovery of Breast Cancer-Related Genes

Current Bioinformatics, 2016, Vol. 11, No. 1


