Searching for essential genes and drug discovery in breast cancer and periodontitis via text mining and bioinformatics analysis

Lan Luo, Weijie Zheng, Chuang Chen and Shengrong Sun

The primary purpose of the study was (1) to search for the essential genes associated with breast cancer and periodontitis, and (2) to identify candidate drugs targeted to these genes for expanding the potential drug indications. The genes related to both breast cancer and periodontitis were determined by text mining. Gene ontology and Kyoto Encyclopedia of Genes and Genomes analysis were performed on these genes, and protein–protein interaction analysis was carried out to export significant module genes. Drug–gene interaction database was employed for potential drug discovery. We identified 221 genes common to both breast cancer and periodontitis. The top six significant enrichment terms and 15 enriched signal pathways were selected. Among 24 significant genes demonstrated as a gene cluster, we found SERPINA1 and TF were significantly related to poor overall survival between the relatively high and low groups in patients. Using the final two genes, 12 drugs were identified that had potential therapeutic effects. SERPINA1 and TF were screened out as essential genes related to both breast cancer and periodontitis, targeting 12 candidate drugs that may expand drug indications. Drug discovery using text mining and analysis of different databases can promote the identification of existing drugs that have the potential of administration to improve treatment in breast cancer. Anti-Cancer Drugs 32: 1038–1045 Copyright © 2021 The Author(s). Published by Wolters Kluwer Health, Inc.

Introduction
Breast cancer is the most common malignant tumor in women, accounting for 30% of female cancers in 2020 [1]. Several factors are known to affect women’s risk of breast cancer, including lifestyle, age, radiation exposure, obesity and so on [2,3]. Besides, recent studies have shown that the human microbiome, including gut microbiome, may be associated with the development of breast cancer [4]. As one of the therapeutic backbones, chemotherapy can be given before or after surgery, which is able to alleviate clinical symptoms, eliminate any metastatic lesion that may exist and improve the survival rate of patients [5,6]. However, cytostatic drugs have various side effects, one of which is oral complications, such as mucositis, xerostomia and periodontal disease, to name a few [7]. In addition, the use of selective estrogen modulators, such as tamoxifen and aromatase inhibitors, have been reported to be associated with significant loss of bone mineral density (BMD), which is cited as a risk factor for periodontitis [8,9].

Periodontitis is a complex disease caused by the growth of aggressive micro-organisms on the teeth [10], and these pathogenic micro-organisms may participate in carcinogenesis through inflammation and immune regulation [11]. A recent study observed that periodontal inflammation plays a role in the early step metastasis of breast cancer by pyroptosis-induced IL-1 beta generation and downstream signaling pathways [12]. In addition, some specific pathogenic bacteria, such as Fusobacterium nucleatum, have been found to be able to enrich in tumors, such as human colorectal cancer (CRC) [13] and breast cancer [14], which are capable of promoting the proliferation of cancer cells and increasing the possibility of metastasis through various mechanisms [15,16], and may involve in tumor recurrence and chemotherapy resistance [17]. Therefore, effective periodontal treatment may be a valuable therapy to prevent breast cancer [18].

Text mining method has become a powerful means to extract information and knowledge from a large number of scientific literature, which can provide clues for the repositioning of existing drugs and help for effective
In this study, we attempted to search for the essential genes related to both breast cancer and periodontitis and expand new indications of exited drugs targeted to these genes, in order to identify new targets and open up new possible approaches to breast cancer treatment. First, we used text mining to obtain the common genes between breast cancer and periodontitis. The in-depth information of these genes was integrated in the manner of functional analysis and enrichment. Next, protein and protein interactions (PPI) analysis was performed to mine public databases and identify significant module genes with more interactions. Lastly, drug–gene interaction analysis was carried out in the drug–gene interaction database (DGIdb), from which the candidate drugs we needed were finally derived. Figure 1 shows the data mining strategy of this study.

Material and methods

Text mining

We used the database pubmed2ensembl (http://pubmed2ensembl.ls.manchester.ac.uk) to perform text mining. Pubmed2ensembl is an open website that makes it easier for biologists to find relevant literature on specific genomic regions or functional gene sets [22]. We input the search terms ‘breast cancer’ and ‘periodontitis’ to obtain a series of genes, and then the common genes were collected for further analysis.

Gene ontology and pathway enrichment

The gene ontology (GO) analysis is always used as a universal and useful approach for annotating genes products and characteristics. The Kyoto Encyclopedia of Genes and Genomes (KEGG) is another frequently-used open database integrating genomic, chemical, and systemic functional information [23]. In this study, we used DAVID, a web-based analysis tool that provides a comprehensive set of applicable annotation services to perform the GO and KEGG enrichment of the common genes.

Protein–protein interactions and module analysis

The protein–protein interactions were analyzed via the online STRING database (version 11.0). First, we input the common genes obtained from the previous steps; the minimum required interaction score was set to the highest confidence (score 0.900). The Cytoscape software was then employed to analyze the PPI works with the TSV file downloaded from the STRING database. We chose the app STRING and Molecular Complex Detection (MCODE) to calculate and select the key gene modules that interconnected clusters in the PPI network. The target genes were collected for the next step.

Gene data analysis

The Gene Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn/index.html) is a developed website that contained a considerable quantity...
of RNA sequencing expression data from the TCGA and the GTEx projects [24]. We conducted differential expression analysis and patient survival analysis of target genes through GEPIA and obtained genes with significant differences in expression in the above two analyses for further study.

Drug–gene interactions
We used DGIdb (http://www.dgidb.org), a valuable database that provides free services for searching the information on drug–gene interactions, to explore the potential drugs or compounds for new treatment. The candidate drugs obtained from the previous step genes may be
the new solutions in treating breast cancer patients with periodontitis.

**Results**

**Text mining, gene ontology and pathway enrichment analysis**

Referring to the text mining analysis described above, we found 769 genes related to breast cancer, 737 genes related to periodontitis and 221 common genes in two lists (Fig. 1). The DAVID website was then employed to perform the GO and KEGG enrichment of the common genes, and Fig. 2a showed the top six significant enrichment terms for BP, CC and MF. The most enriched BP annotations included: response to organic substance, regulation of cell proliferation and cellular response to a chemical stimulus. As for CC annotations, it was significantly enriched in extracellular space, cell surface and extracellular region part. In MF category, the following items accumulated the most: receptor binding, cytokine activity and cytokine receptor binding. Next, the enriched KEGG signal pathway annotations were analyzed and 15 pathways were selected, such as Cytokine–cytokine receptor interaction, TNF signaling pathway and identical protein binding (Fig. 2b).

**Protein–protein interactions and module analysis**

The protein–protein interactions were illustrated using the STRING website. The common genes obtained from above were all input, and STRING visualized the interactions between them (Supplementary Figure S1, Supplemental digital content 1, http://links.lww.com/ACD/A393). The data in TSV format was imported into the Cytoscape software, where the significant gene modules were demonstrated using the app STRING and MCODE. The analysis showed that Cluster 1 (score: 14.87) contained 24 nodes and 171 edges (Fig. 3), and Cluster 2 (score: 7.75) contained 17 nodes and 62 edges (Supplementary Figure S2, Supplemental digital content 1, http://links.lww.com/ACD/A393). We selected the genes involved in Cluster 1 to participate in the gene data analysis.

**Gene data analysis**

GEPIA showed the expression data of the 24 key genes collected from the previous step. We generated the differential expression analysis and patient survival analysis of these genes and the results were presented as boxplots and survival plots. The expression of IGF1, ALB, CYR61, SPP1, CSF1, IGFBP3, TIMP1, SERPINA1, TF, SPARC and IL6 was significantly different between the normal and tumor groups, while the overall survival time results suggested that the expression of SERPINA1 and TF between the relatively high and low group in patients was significantly related to poor overall survival (Fig. 4a,b, Supplementary Figure S3–S11, Supplemental digital content 1, http://links.lww.com/ACD/A393). As a result, we could choose SERPINA1 and TF for drug–gene interactions analysis.

**Drug–gene interactions analysis**

Using the final two genes, SERPINA1 and TF, we identified 12 drugs that had potential therapeutic effects for periodontitis in breast cancer patients, including anticancer drugs, anti-inflammatory drugs, astringent, cholinesterase inhibitors and so on (Fig. 5 and Supplementary Table S1, Supplemental digital content 2, http://links.lww.com/ACD/A394). Four drugs related to SERPINA1, six drugs related to TF, and two drugs related to both genes. Some metal-based anticancer compounds had been screened out among these candidate drugs, such as Cisplatin, a platinum-based chemotherapy drug used to treat various types of cancers [25]; topical medications were also involved, such as aluminum acetate, which is an external astringent which tends to coagulate skin protein, reduce fluids exudation and promote the regression of inflammation.

**Discussion**

Although many studies have been carried out on the exact etiology and pathogenesis of breast cancer, it is still not completely clear [26]. As one of the factors that affect the prognosis quality of life of breast cancer
patients, periodontitis has been proved to be related to the occurrence and development of breast cancer \[27–29\]. Here, we identified two key genes, namely \textit{SERPINA1} and \textit{TF}, associated with both breast cancer and periodontitis using text mining and gene data analysis and target 12 drugs found to have potential therapeutic effects.

Nowadays, treatment for periodontitis is generally centered on removing supragingival plaque and calculus, supplemented by antimicrobial therapy and some personalized oral hygiene instructions when necessary \[30\]. However, among breast cancer survivors, the cancer diagnosis has diverted people’s attention to prolonging survival time and improving quality of life, so the number of

The expression and the overall survival time of \textit{SERPINA1} and \textit{TF}. (a) The expression of \textit{SERPINA1} and \textit{TF} was significantly different between the normal and tumor groups. (b) The low expression group of \textit{SERPINA1} and \textit{TF} in patients was significantly related to poor overall survival \(P < 0.005\).
patients who have access to preventive services such as BMD measurement and oral health care is less than that of the age-matched control group [8]. Therefore, developing effective new drugs seems to be extremely important.

The SERPINA1 gene is located on the long arm of chromosome 14 (14q32.1) [31]. It encodes α1-antitrypsin (ATT), the primary protease inhibitor (PI) in human serum, and the mutations in SERPINA1 genes cause alpha-1-antitrypsin deficiency (AATD), which manifests as pulmonary emphysema and liver cirrhosis [32]. Recently, SERPINA1 has been identified as a biomarker for a variety of diseases, such as papillary thyroid carcinoma [33], non–small-cell lung cancer [34] and breast cancer [35]. Chan et al. [36] found the single gene SERPINA1 is a valuable predictor of survival in ER+ and ER+/HER2+ breast cancer patients, high expression of SERPINA1 predicted better clinical outcomes in ER+ and ER+/HER2+ patients. Du et al. [37] analyzed SERPINA1 as one of the autophagy-related genes that had significant prognostic values for breast cancer. Tissue factor (TF) is a 47 kDa transmembrane glycoprotein compounded with factor VII for procoagulant function during human blood coagulation [35]. TF expresses in a wide range of tumors, including breast cancer, of which high expression is associated with cancer metastasis and overall survival. Overexpression of TF leads to a hypercoagulable state, and thromboembolism is among the common complications in cancer patients [38]. Meanwhile, the complex TF-fVIIa may stimulate the expression of numerous malignant phenotypes in breast cancer cells [39]. Therefore, TF turns out to be a potentially attractive target for potential antiangiogenic and platelet adhesion blocking strategy [40,41]. Gingival crevicular fluid (GCF) is a vital target in the search of both periodontitis and gingivitis biomarkers. It has been demonstrated that the protein composition may reflect the pathophysiology of periodontitis using proteomics approach [42]. Preiano et al. [43] identified the significant peptide signatures in GCF between two groups of gingivitis and healthy subjects by performing MALDI-TOF MS comparative analysis, and they found five peptides differentially expressed, including the C-terminal fragment of AAT. In a word,

**Fig. 5**

Twelve drugs were collected that had potential therapeutic effects for periodontitis in breast cancer patients. Using the final list of two genes as the potential targets in the drug–gene interaction analysis, the list of drugs was selected as possible drug treatments.
the investigations above supply a basis for targeting the key genes SERPINA1 and TF for potential treatment strategies.

According to our study, 12 drugs were screened out that may act in a variety of ways to achieve potential therapeutic effects for periodontitis in breast cancer patients. Bacterial infection induces disruption of a range of signaling pathways, including inflammatory pathways, and the drugs capable of interfering with these pathways may be therapeutic. Guan et al. [44] reported that the NLRP3/Caspase-1/IL-1 beta signaling pathway participated in the estrogen-mediated periodontitis and the consequent bone loss. In the meantime, the NLRP3 inflammasome and its downstream IL-1 beta pathways have been confirmed to promote tumor growth and metastasis in animal and human breast cancer models [45], and they are considered as a novel potential target and could open new frontiers in breast cancer treatment [46]. Several NLRP3 inflammasome activation inhibitors were patented, and the enormous effort to discover and develop activation inhibitors of NLRP3 inflammasome may provide new insights into the treatment of inflammatory protein regulatory pathways. PTGS2 is another key gene in inflammation, which was highly regulated in both periodontitis and rheumatoid arthritis [47]. It has been proved to be associated with breast cancer risk [48,49]. Tian et al. [50] described the potential prognostic value for COX-2 in basal-like breast cancer, the high COX-2 expression levels related to poor outcomes. Intervention with these inflammatory factors and signaling pathways may provide new ideas for future treatment.

There existed some limitations in our study associated with the databases we used and the keywords we chose in the initial step. These databases are in a continuously updated state, and the threshold score we set are likely to be subjective. Although these drugs could be a new perspective for effective treatment, it is essential to execute further clinical trials to confirm their functions and new indications.

**Conclusion**

In summary, we identified SERPINA1 and TF as key candidate genes related to both breast cancer and periodontitis and screened out 12 drugs targeting the candidate genes by using text mining and analysis of different databases. Efficacious oral hygiene care, combined with the application of new indications for exited drugs, might bring great benefits to the occurrence and development of breast cancer.

**Acknowledgements**

The authors wish to acknowledge all researchers who provided the data. Thanks are also due to Bin Zhao and Yuting Chen for valuable discussion.

This work was supported by grants from the Fundamental Research Funds for the Central Universities (2042019kf0229).

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70:7–30.
2. Paula NCNM, Martins JAM, Amaral LM, Rhana P, Tavares EC, Leite WS, et al. Breast cancer: is grief a risk factor? Rev Assoc Med Bras (1992) 2018; 64:595–600.
3. Ponzoni M, Breast cancer prevention: can women’s expectations be met? Oncologist 2016; 21:2–3.
4. Fernández MF, Reina-Pérez I, Astorga JM, Rodríguez-Carrillo A, Plaza-Diaz J, Fontana L. Breast cancer and its relationship with the microbiota. Int J Environ Res Public Health 2018; 15:1747.
5. Akram M, Iqbal M, Danialy M, Khan AU. Awareness and current knowledge of breast cancer. Biol Res 2017; 50:33.
6. Barbeck N, Gnant M. Breast cancer. Lancet 2017; 389:1134–1150.
7. Willershausen I, Schmidtmann I, Azapoor A, Kledtke J, Willershausen B, Hasenburg A. Association between breast cancer chemotherapy, oral health, and chronic dental infections: a pilot study. Odontolontology 2019; 107:401–408.
8. Taichman LS, Griggs JJ, Inglehart MR. Periodontal disease, perceived oral health, and dental care utilization among breast cancer survivors. J Public Health Dent 2015; 75:148–156.
9. Taichman LS, Inglehart MR, Giannobile WV, Braun T, Kolenic G, Van Poznak C. Periodontal Health in Women With Early-Stage Postmenopausal Breast Cancer Newy on Aromatase Inhibitors: A Pilot Study. J Periodontol 2015; 86:906–916.
10. Mombelli A. Microbial colonization of the periodontal pocket and its significance for periodontal therapy. Periodontal 2000 2000; 18:85–96.
11. Chung PC, Chan TC. Association between periodontitis and all-cause and cancer mortality: retrospective elderly community cohort study. BMC Oral Health 2020; 20:168.
12. Cheng R, Biller S, Liu C, Haldar S, Choudhury D, Tripathi M, et al. Periodontal inflammation recruits distant metastatic breast cancer cells by increasing myeloid-derived suppressor cells. Oncogene 2020; 39:1543–1556.
13. Brennan CA, Garrett WS. Fusobacterium nucleatum - symbiotic, opportunist and oncobacter. Nat Rev Microbiol 2019; 17:156–166.
14. Parhi L, Alon-Maimon T, Sol A, Nejman D, Shhadeh A, Fainsod-Levi T, et al. Breast cancer colonization by Fusobacterium nucleatum accelerates tumor growth and metastatic progression. Nat Commun 2020; 11:3259.
15. Guo P, Tian Z, Kong X, Yang L, Shan X, Dong B, et al. FadA promotes DNA damage and progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation of chk2. J Exp Clin Cancer Res 2020; 39:202.
16. Yu MR, Kim HJ, Park HR. Fusobacterium nucleatum accelerates the progression of colitis-associated colorectal cancer by promoting EMT. Cancers (Basel) 2020; 12:E2728.
17. Xu T, Guo F, Yu Y, Sun T, Ma D, Han J, et al. Fusobacterium nucleatum promotes chemoresistance to colorectal cancer by modulating autophagy. Cell 2017; 170:548–563.e16.
18. Shao J, Wu L, Leng WD, Fang C, Zhu YJ, Yin HY, et al. Periodontal disease and breast cancer: a meta-analysis of 1,73,162 participants. Front Oncol 2018; 8:9.
19. Hsiao YY, Lu TP. Text-mining in cancer research may help identify effective treatments. Transl Lung Cancer Res 2019; 8:5460–5463.
20. Zhang N, Xu W, Wang S, Qiao Y, Zhang X. Computational drug discovery in chemotherapy-induced alopecia via text mining and biomedical databases. Clin Ther 2019; 41:972–980.e8.
21. Kirk J, Shah N, Noll B, Stevens CB, Lawler M, Mougeot FB, Mougeot JC. Text mining-based in silico drug discovery in oral mucositis caused by high-dose cancer therapy. Support Care Cancer 2018; 26:2695–2705.
22. Baran J, Gernert M, Haussersler M, Menadic G, Bergman CM, pubmed2ensembl: a resource for mining the biological literature on genes. PLoS One 2011; 6:e24716.
23. Kanaihisa M, Goto S, Yato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012; 40:D109–D114.
Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res 2019; 47:W98–W102.

Ndagi U, Mhlongo N, Soliman ME. Metal complexes in cancer therapy - an update from drug design perspective. Drug Des Devel Ther 2017; 11:599–616.

Parida S, Sharma D. The power of small changes: Comprehensive analyses of microbial dysbiosis in breast cancer. Biochim Biophys Acta Rev Cancer 2019; 1871:392–405.

Sfreddo CS, Maier J, De David SC, Susin C, Moreira CHC. Periodontitis and breast cancer: a case-control study. Community Dent Oral Epidemiol 2017; 45:545–551.

Shi T, Min M, Sun C, Zhang Y, Liang M, Sun Y. Periodontal disease and susceptibility to breast cancer: a meta-analysis of observational studies. J Clin Periodontol 2018; 45:1025–1033.

Dizdar O, Hayran M, Guven DC, Yılmaz TB, Taheri S, Akman AC, et al. Increased cancer risk in patients with periodontitis. Curr Med Res Opin 2017; 33:2195–2200.

Graziani F, Karapetsa D, Alonso B, Herrera D. Nonsurgical and surgical treatment of periodontitis: how many options for one disease? Periodontal 2000 2017; 78:152–188.

DeLuca DS, Poluzioroviene E, Taminskiene V, Wrenger S, Utkus A, Valiulis et al. Increased cancer risk in patients with periodontitis. Curr Med Res Opin 2020; 36:599–616.

Qu T, Li YP, Li XH, Chen Y. Identification of potential biomarkers and drugs for papillary thyroid cancer based on gene expression profile analysis. Mol Med Rep 2016; 14:5041–5048.

Falanga A, Szczepek A, Russo D. Cancer tissue procoagulant mechanisms and the hypercoagulable state of patients with cancer. Semin Thromb Hemost 2015; 41:756–764.

Koizume S, Miyagi Y. Breast cancer phenotypes regulated by tissue factor-factor VII pathway: possible therapeutic targets. World J Clin Oncol 2014; 5:908–920.

Liu S, Zhang Y, Zhao X, Wang J, Di C, Zhao Y, et al. Tumor-specific silencing of tissue factor suppresses metastasis and prevents cancer-associated hypercoagulability. Nano Lett 2019; 19:4721–4730.

Ruf W, Yokota N, Schaffner F. Tissue factor in cancer progression and angiogenesis. Thromb Res 2010; 125 (Suppl 2):S36–S38.

Preiànò M, Savino R, Villeva C, Pelaia C, Terracciano R. Gingival crevicular fluid peptidome profiling in healthy and in periodontal diseases. Int J Mol Sci 2020; 21:E9270.

Quagliariello V, Bonelli A, Caronna A, Lombari C, Grimaldi I, Conforti G, et al. Hyperglycemia increased nivolumab-induced cardiotoxicity, enhanced immunoresistance of ER alpha+, PR+, HER2-breast cancer cells modulating growth factors and NLRP3 expression. Ann Oncol 2020; 31:S1097.

Nolani R, Agrawal N, Ansari M, Jain S, Jindal A. COX-2 expression in breast carcinoma with correlation to clinicopathological parameters. Asian Pac J Cancer Prev 2018; 19:1971–1975.

Zhuang X, Zhang H, Li X, Li X, Cong M, Peng F, et al. Differential effects on lung and bone metastasis of breast cancer by Wnt signalling inhibitor DKK1. Nat Cell Biol 2017; 19:1274–1285.

Copyright © 2021 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.