Original Article

Identification of the Disrupted Pathways Associated with Periodontitis Based on Human Pathway Network

Wen Sun¹, Lin Han², Wenmao Xu³, Yazhen Sun⁴

¹Beijing springer Medical Research Institute, Beijing, China; ²Department of Medical, Medical editorial department, ⁴Department of Planning, Shandong Science and Technology Press, Jinan, China; ³Jinan EBM science and technology development center, Jinan, China

Keywords

periodontitis; disrupted pathways; protein-protein interactions; EB co-expression

Correspondence

Yazhen Sun,
E-mail: pxf0531@163.com

DOI: 10.1515/ii-2017-0143

Abstract

Objective: The objective of this work is to search for a novel method to explore the disrupted pathways associated with periodontitis (PD) based on the network level. **Methods:** Firstly, the differential expression genes (DEGs) between PD patients and cognitively normal subjects were inferred based on LIMMA package. Then, the protein-protein interactions (PPI) in each pathway were explored by Empirical Bayesian (EB) co-expression program. Specifically, we determined the 100th weight value as the threshold value of the disrupted pathways of PPI by constructing the randomly model and confirmed the weight value of each pathway. Meanwhile, we dissected the disrupted pathways under the weight value > the threshold value. Pathways enrichment analyses of DEGs were carried out based on Expression Analysis Systematic Explored (EASE) test. Finally, the better method was selected based on the more rich and significant obtained pathways by comparing the two methods.

Results: After the calculation of LIMMA package, we estimated 524 DEGs in all. Then we determined 0.115222 as the threshold value of the disrupted pathways of PPI. When the weight value > 0.115222, there were 258 disrupted pathways of PPI enriched in. Additionally, we observed those 524 DEGs that were enriched in 4 pathways under EASE=0.1.

Conclusion: We proposed a novel network method inferring the disrupted pathway for PD. The disrupted pathways might be underlying biomarkers for treatment associated with PD.

Introduction

Periodontitis (PD) is a chronic inflammatory disease, resulting in loss of connective tissue and alveolar bone support of the teeth^[1]. It is the most frequent cause of tooth loss in the adult^[2]. In the early stages, PD has very few symptoms, it may include: redness or bleeding of gums while brushing teeth, gum swelling, loose teeth in the later stages, *et al.* PD is a multi-factorial disease and as such, the significant elements include not only the presence of pathogenic bacteria and the immune mechanism, but also the genetic predisposition^[3]. Though the mechanism of PD was not very clear, it has been evidenced by some authors that there is a joint influence of polymorphisms in multiple genes^[4], such as IL-10^[5], IL-6^[6] and SOCS1– 820 polymorphism^[7].

However, the complete underlying genomic or gene expression variation that contributes to the presence of PD is still unknown. Network-based approaches especially co-expression network offer a more effective mean to dig potential malignancy diagnostic molecular based on connecting them together. Co-expression networks is gradually used to study the disease mechanisms^[8] and provides the systems level view of dysregulated pathways^[9]. The basic premise of co-expression analysis is that strongly correlated genes are likely to be functionally associated. Further we can gain a clear insight into the important genes and pathways of a variety of diseases, many of which are applicable to the early detection and treatment in cases.

In this research, we downloaded all pathways and chosen the pathways for further research which the number of intersections genes between gene expression profile with

each pathway >5 and pathway and the ratio with gene number in each pathway was greater than 0.5. Co-expression analysis was conducted for the genes in these pathways, and its weight value was calculated. Building randomly a model, we kept the 100th weight value as a threshold and screened the pathways as the disrupted pathways with the weight value > the threshold. Meanwhile, we performed the pathways enrichment analysis for the DEGs. In comparison with individual gene-based approaches, we acquired more accurate results in pathway enrichment analysis of the co-expression of genes from pathways. We expected that pathway network approach would provide new opportunities for uncovering disrupted pathways caused by disease. Moreover, discriminatory pathways between patients and cognitively normal subjects may facilitate the interpretation of functional alterations during disease progression.

Materials and methods

Data collection and preprocessing

The microarray expression profile of E-GEOD-16134^[10] was recruited from Array Express database (<http://www.ebi.ac.uk/arrayexpress/>). A total of 310 samples (241 PD cases and 69 controls) were collected on the platform of Affymetrix Gene Chip Human Genome U133 Plus 2.0 in the present study.

In order to reduce the influence of nonspecific factors about the datasets, we conducted the preprocessing for all the original expression information before the analysis included of the background correction and the normalization using robust multichip average (RMA) method^[11] and quantile based algorithm^[12], respectively. Perfect match and mismatch value were revised and selected via Micro Array Suite 5.0 (MAS 5.0) algorithm^[13] and the median method, respectively. The following the data were screened by feature filter method of gene filter package. Each probe was mapped to one gene by annotate package, and the probe is discarded if it can't match any genes.

Detection of DEGs

At the present study, the DEGs between PD patients and control subjects were screened by the linear models for microarray data (LIMMA) package^[14]. T-test and F-test were carried on the matrix, and then the p-values were transformed to -log₁₀. Empirical Bayes (EB)^[15] statistics and a false discovery rate (FDR)^[16] calibration of P-values for

the data were conducted by lmFit function. The DEGs were selected from linear after inspection which needed to meet the following cut-off criteria: values of |log fold change(FC)| ≥ 2, p-values < 0.05.

Pathway enrichment analysis for the DEGs

To further investigate the enriched pathways of the DEGs of PD, a pathway analysis was performed based on the Reactome (<http://www.reactome.org>), which is a manually curated open-source open-data resource of human pathways and reactions. Reactome aims to systematically associate human proteins with their molecular and cellular functions in order to create a knowledge base of human biological reactions, pathways and processes that can be used both as an online encyclopedia and as a systems biology platform for data mining and analysis^[17-19]. We imported the DEGs to the online tool of Database for Annotation, Visualization and Integrated Discovery^[20] (DAVID, <http://david.abcc.ncifcrf.gov/tools.jsp>), and obtained all pathways these genes enriched. We got the enrichment pathways according to the EASE for 0.1.

Pathway enrichment analysis based on EB co-expression networks

We downloaded all human pathways (1675) from the reactome database (<http://www.reactome.org>) which was a collaboration among groups, to develop an open source curated bioinformatics database of human pathways and reactions^[17]. Meanwhile, the pathways without genes or only one gene were removed. In this context, this study obtained 1639 pathways in all for further research.

We calculated the average gene number G of each pathway in these 1639 pathways. The formula was as followed: $G = \frac{\text{Total number of genes in all pathways}}{\text{all pathways}}$. We defined the number of genes in each pathway as A, the intersection number of genes between each pathway with the expression profile as B. as a consequence, we selected 1314 pathways used in this method under $B > 5$ as well as $B/A > 0.5$.

Determining the threshold of disrupted pathways

It is worth noting a few other approaches that have been developed for co-expression analysis. EB approach^[21] is proven to be a useful complement to existing differential expression methods by simulations and case studies respectively. The approach provides an FDR controlled list of interesting pairs along with pair-specific posterior

probabilities that can be used to identify particular types of DC. In order to determine how to choose the disrupted pathways for further research, here we built one model randomly. G genes were selected randomly from the expression profile. In this section, we conducted the EB co-expression analysis between the genes in the 1314 pathways. The co-expression relationships pairs in each pathway was symbolized as C . In the article, we screened the co-expression relationships pairs D under $FDR \leq 0.05$. Meanwhile, we defined the D/C as the weight value of each pathway. A random sample of 10000 times was performed, and the corresponding 10000 weight value were obtained. These weight value were ranked in descending order, 100th P -value was defined as the threshold ($FDR = 100/10000 = 0.01$).

To sum up, the disrupted pathways were finally obtained based on the weight value greater than the threshold.

In-depth analyses of disrupted pathways

In order to further clarify the specific differences between disrupted pathways, gene compositions of the modules were analyzed and compared. Here we observed that whether there existed some DEGs in disrupted pathways. Then, statistical analysis was conducted on these DEGs that appeared in the disrupted pathways. The number of DEGs was more in the disrupted pathways, and the pathways meant to be more associated with the disease. Meanwhile, we picked two disrupted pathways randomly, constructed

the co-expression network and observed the interaction relationship among genes in these pathways.

Results

Detection of DEGs

In the present study, a total of 20107 genes of the dataset associated with PD were preprocessed to identify DEGs using LIMMA package. Finally, there were 524 DEGs were detected between PD patients and controls under the criteria of $|\log FC| \geq 2$ and $P < 0.05$.

Pathways enrichment analysis for the DEGs

In this paper, for KEGG pathway enrichment analysis, the enriched pathways of the DEGs showed that Signaling in Immune system ($P = 9.78E-10$), Hemostasis ($P = 1.41E-04$), Integrin cell surface interactions ($P = 6.77E-03$) and Metabolism of vitamins and cofactors ($P = 9.07E-02$) were significant pathways with EASE for 0.1.

Pathway enrichment analysis based on EB co-expression networks

We calculated the average gene number 44 of each pathway in these 1639 pathways. Moreover we selected 1314 pathways based on the intersection number of genes > 5 between each pathway with the expression profile as well as the ratio of the intersection number of genes and the number of genes in each pathway > 0.5 . Finally, 258 disrupted pathways were finally obtained based on the weight value greater than the threshold of 0.115.

In-depth analyses of disrupted pathways

Observing the DEGs in disrupted pathways, we found 249 enriched pathways in all. Removing the pathways without DEGs, a total of 75 disrupted pathways were selected. Meanwhile, under the intersection number of genes between the disrupted pathways and DEGs > 3 , we obtained 6 disrupted pathways (Figure 2). They might be related to PD.

In the furthermore, we obtained the co-expression network in Ubiquitin-dependent degradation of Cyclin D and Viral mRNA Translation based on the EB approach, especially. The result was shown in Figure 3. From this imagine, we could notice that almost genes in each pathway were closely related.

Figure 1. The cluster heat map of the 258 disrupted pathways. The colors represented the pathway weight values. Horizontal axis represents samples; vertical coordinate represents disrupted pathways.

Classical pathways enrichment analysis of the gene expression data set provided little insight into the biological processes associated with the treated PD. What was more, the high number of potential false positives hampered the results and correction for multiple testing eliminated most of these genes from the final list. Network analysis offers some advantages over classical analysis by being able to incorporate additional information from multiple sources^[22-24]. In our work, we applied a similar methodology integrating network information into pathway activity inference rather than single genes activities.

(CLRs) and HIV Infection were more significant.

Infection is the invasion of an organism's body tissues by disease-causing agents and the reaction of host tissues to these organisms and the toxins they produce. Infectious disease, also known as communicable disease, is illness resulting from an infection. Persistent infections cause millions of deaths globally each year. For example, infectious diseases resulted in 9.2 million deaths in 2013^[25]. Not all infections are symptomatic^[26]. In certain cases, infectious

diseases may be asymptomatic for much or even all of their course in a given host. In the latter case, the disease may only be defined as a "disease" in hosts who secondarily become ill after contact with an asymptomatic carrier. Techniques like hand washing and wearing face masks can help prevent infections from being passed from the surgeon to the patient or vice versa. Frequent hand washing remains the most important defense against the spread of unwanted organisms. C-type lectin receptors (CLRs) are a large family of receptors that encompass upwards of 1000 members with diverse functions including cell adhesion, complement activation, tissue remodeling, platelet activation, endocytosis, phagocytosis, and activation of innate immunity^[27,28]. CLRs are important for cell–cell communication and host defense against pathogens through the recognition of specific carbohydrate structures. In addition, there may also be a direct activation of acquired immunity. Yan *et al.*^[29] overview the current knowledge of CLRs signaling and the application of their ligands on tumor-associating immune response and realize the specific regulation of CLRs signaling by modulating tumor micro environment should lead to the best application of CLRs biology. Here we believed that Infectious disease and C-type lectin receptors will have a certain impact on PD.

Our approach provided enhanced diagnostic power achieving higher accuracy, sensitivity and specificity. These disrupted pathways have pivotal roles in triggering the development of PD.

Declarations

Acknowledgements

No.

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

W Sun and L Han designed the experiments, performed the experiments and drafted the manuscript; WM Xu performed the experiments; YZ Sun revised the manuscript. All authors read and approved the final manuscript.

References

- Pihlstrom BL, Michalowicz BS, Johnson NW. Periodontal diseases. *The Lancet*. 2005; 366(9499): 1809-20.
- Page RC. Milestones in periodontal research and the remaining critical issues. *Journal of Periodontal Research*. 1999; 34(7): 331-9.
- Borrell LN, Papapanou PN. Analytical epidemiology of periodontitis. *Journal of Clinical Periodontology*. 2005;32(s6): 132-58.
- Loos BG, John RP, Laine ML. Identification of genetic risk factors for periodontitis and possible mechanisms of action. *Journal of clinical periodontology*. 2005; 32(s6): 159-79.
- Armingohar Z, Jørgensen JJ, Kristoffersen AK, Schenck K, Dembic Z. Polymorphisms in the interleukin-10 gene and chronic periodontitis in patients with atherosclerotic and aortic aneurysmal vascular diseases. *Journal of oral microbiology*. 2015;7.
- Gabriela Teixeira F, Mendonça SA, Menezes Oliveira K, Barbosa dos Santos D, Miranda Marques L, Mendonça Amorim M, *et al.* Interleukin-6 c.-174G> C Polymorphism and Periodontitis in a Brazilian Population. *Molecular biology international*. 2014;2014.
- Guedes RA, Planello AC, Andia DC, De Oliveira NF, de Souza AP. Association of SOCS1– 820 (rs33977706) gene polymorphism with chronic periodontitis: A case–control study in Brazilians. *Meta gene*. 2015;5:124-8.
- Nayak RR, Kearns M, Spielman RS, Cheung VG. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Res*. 2009 Nov;19(11):1953-62.
- Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 2011 Jun 16;474(7351):380-4.
- Papapanou PN, Behle JH, Kebschull M, Celenti R, Wolf DL, Handfield M, *et al.* Subgingival bacterial colonization profiles correlate with gingival tissue gene expression. *BMC microbiology*. 2009;9(1):221.
- Ma L, Robinson LN, Towle HC. ChREBP*MLx is the principal mediator of glucose-induced gene expression in the liver. *J Biol Chem*. 2006 Sep 29;281(39):28721-30.
- Rifai N, Ridker PM. Proposed cardiovascular risk assessment algorithm using high-sensitivity C-reactive protein and lipid screening. *Clinical chemistry*. 2001;47(1):28-30.
- Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ. The utility of MASS expression summary and detection call algorithms. *BMC Bioinformatics*. 2007;8:273.
- Smyth G, Thorne N, Wettenhall J. LIMMA: Linear Models for Microarray Data User's Guide, 2003. URL <http://www.bioconductor.org>. 2005.
- Datta S, Satten GA, Benos DJ, Xia J, Heslin MJ, Datta S. An empirical bayes adjustment to increase the sensitivity of detecting differentially expressed genes in microarray experiments. *Bioinformatics*. 2004;20(2):235-42.
- Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics*.

- 2003;19(3):368-75.
- 17 Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D691-7.
- 18 Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D619-22.
- 19 Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 2005 Jan 1;33(Database issue):D428-32.
- 20 Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols.* 2008;4(1):44-57.
- 21 Dawson JA, Kendziorski C. An Empirical Bayesian Approach for Identifying Differential Coexpression in High-Throughput Experiments. *Biometrics.* 2012;68(2):455-65.
- 22 Diez D, Wheelock ÅM, Goto S, Haeggström JZ, Paulsson-Berne G, Hansson GK, *et al.* The use of network analyses for elucidating mechanisms in cardiovascular disease. *Molecular BioSystems.* 2010;6(2):289-304.
- 23 Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics.* 2011;12(1):56-68.
- 24 Lee D-S, Park J, Kay K, Christakis N, Oltvai Z, Barabási A-L. The implications of human metabolic network topology for disease comorbidity. *Proceedings of the National Academy of Sciences.* 2008;105(29):9880-5.
- 25 Wang H, Liddell CA, Coates MM, Mooney MD, Levitz CE, Schumacher AE, *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet.* 2014;384(9947):957-79.
- 26 Ljubicin-Sternak S, Meštrović T. Chlamydia trachomatis and Genital Mycoplasmas: Pathogens with an Impact on Human Reproductive Health. *Journal of pathogens.* 2014;2014.
- 27 Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. *Immunological reviews.* 1998;163(1):19-34.
- 28 Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. *Febs Journal.* 2005;272(24):6179-217.
- 29 Yan H, Kamiya T, Suabjakyong P, Tsuji NM. Targeting C-type lectin receptors for cancer immunity. *Frontiers in immunology.* 2015;6.