Behavioural Brain Research 320 (2017) 136–142

Contents lists available at ScienceDirect

Behavioural Brain Research

journal homepage: www.elsevier.com/locate/bbr

Research report

Using the BITOLA system to identify candidate molecules in the interaction between oral lichen planus and depression☆

Yuanbo Zhan[a,1], Shuang Zhou[a,1], Ying Li[a,1], Sen Mu[a,1], Ruijie Zhang[b], Xuejing Song[c], Feng Lin[a], Ruimin Zhang[d,*], Bin Zhang[a,e,**]

[a] Institute of Hard Tissue Development and Regeneration, The Second Affiliated Hospital of Harbin Medical University, Harbin 150001, Heilongjiang, China
[b] Colleges of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China
[c] Department of Stomatology, The Central Hospital of Liaoyang City, Liaoyang 111000, China
[d] Department of Periodontology and Oral Mucosa, The Second Affiliated Hospital of Harbin Medical University, Harbin, 150086, China
[e] Heilongjiang Academy of Medical Sciences, Harbin 150001, Heilongjiang, China

HIGHLIGHTS

- NCAM1 and CD4 were differently expressed in both oral lichen planus and depression patients when compared to controls.
- The expressions of NCAM1 and CD4 were observed both in oral and brain.
- NCAM1 and CD4 was identified as involved or potentially involved in the interaction between OLP and depression.
- Using the text mining can offer a new clue for the experimenters and for developing innovative therapeutic strategies.

ARTICLE INFO

Article history:
Received 28 August 2016
Received in revised form
19 November 2016
Accepted 28 November 2016
Available online 29 November 2016

Keywords:
Mouth diseases
Mental disorders
Text mining
Gene expression profiling

ABSTRACT

Exacerbations of oral lichen planus (OLP) have been linked to the periods of psychological stress, anxiety and depression. The specific mechanism of the interaction is unclear. The aim of this study was to explore the candidate genes or molecules that play important roles in the interaction between OLP and depression. The BITOLA system was used to search all intermediate concepts relevant to the "Gene or Gene Product" for OLP and depression, and the gene expression data and tissue-specific gene data along with manual checking were then employed to filter the intermediate concepts. Finally, two genes (NCAM1, neural cell adhesion molecule 1; CD4, CD4 molecule) passed the follow-up inspection. By using the text mining can formulate a new hypothesis: NCAM1 and CD4 were identified as involved or potentially involved in the interaction between OLP and depression. These results offer a new clue for the experimenters and hold promise for developing innovative therapeutic strategies for these two diseases.

1. Introduction

Oral lichen planus (OLP), a chronic inflammatory disease, is always characterized by bilateral white striations or plaques, usually involving the buccal mucosa, dorsum of the tongue and gingivae [1]. It is found commonly in adults (50–55 years of age) and predominantly affects women usually by a 1.4:1 ratio over men. Although many elements, including viral [2–5], genetic [6], and immunological factors [7] are known to be implicated in the OLP, the precise cause of lichen planus or the etiology is still uncertain. In the past two decades, stress, anxiety and depression have frequently been reported as possible factors related to the development of OLP [8,9]. OLP is one of the dermatological disorders. Due to interaction between skin and mind [10], the interesting area of the interface between psychiatry and dermatology has aroused the attention of many researchers. Some evidence has also showed psychiatric comorbidity influences the development and course of dermatologic diseases via the effects of stress, depression, and anxiety [11]. On the other hand, mental intervention or psychotherapy

☆ This work was performed at Institute of Hard Tissue Development and Regeneration, the Second Affiliated Hospital of Harbin Medical University, Harbin 150001, Heilongjiang, China.

* Corresponding author at: Institute of Hard Tissue Development and Regeneration, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, China.
** Corresponding author.
E-mail addresses: yzzhangrm@126.com (R. Zhang), zhangbinyzz@126.com (B. Zhang).

[1] Join the first author.

http://dx.doi.org/10.1016/j.bbr.2016.11.047
0166-4328/

may produce beneficial effects for OLP. Thus psychiatric evaluation can be considered for OLP patients with routine treatment protocols recommended [10]. However, the underlying molecular mechanism of the interaction is not clear.

With the passage of time, biomedical literature has rapidly expanded. Typically by special keywords, one may obtain much literature relevant to the keywords unless some areas with few or no research; however; there are always a lot of redundancies that need manual checking. Moreover, this process makes it difficult for researchers to absorb all relevant information within a short time in their fields; even more so outside them [12]. Fortunately, sophisticated computer-supported text mining tools have been created. It has also been demonstrated that important connections and valuable hypotheses can be extracted by linking findings across scientific literature with the use of literature mining methods and tools [13].

Generating new discoveries based on different literature, not described in any research paper, was first proposed by Swanson[14,15]. In silico model ABC, suggested by Swanson, performs a search for new indirect relations between two disjoint set of literature [13]. Later, the model has been developed into two main approaches. One is the closed discovery process, and the other is the open discovery process. The former is mainly focused on the test of a starting hypothesis. For example, if there are two starting domains "a" and "c", and two pieces of corresponding literature A and C are retrieved, in the meantime, the common terms "b" is selected based on "b" appears in both pieces of literature [13]. The latter is characterized by the absence of advance specification of target concepts. Hristovski [12] and Weeber [16] et al. provide a review for the detail information about different functions of literature mining tools. Some researchers have used the text mining tools to identify candidate genes for diseases [12,17].

In this study, we used the text mining to discover the candidate genes or molecules that may play important roles in the interaction between OLP and depression. To our knowledge, this is the first time using this strategy in this area.

2. Materials and methods

2.1. Searching the intermediate concepts using BITOLA

BITOLA, one of the most remarkable tools, is an interactive literature-based biomedical discovery support system [18]. The main purpose of this system is to help the researchers make new discoveries by discovering potentially new relations between biomedical concepts [19]. In the BITOLA system, set of concepts currently contains Medical Subject Headings (MeSH), which is utilized to index Medline, and human genes from the Human Genome Organization (HUGO).

The original and the simplest purpose we wanted is to explore which gene or gene products are linked to both OLP and depression, and these gene or gene products may play key roles in the communication between OLP and depression. Specifically, in the BITOLA system, the algorithm [16] for discovering new relations between medical concepts is described in Fig. 1.

According to the proposed instruction in the tool, the closed discovery system was used. Briefly, "Lichen Planus, Oral" was used as starting concept (Unified Medical Language System, semantic types: disease or syndrome), and "Major Depressive Disorder" and "Depressive disorder" were adopted as the end concepts (semantic types: Mental or Behavioral Dysfunction). Last, intermediate concepts were retrieved. In this study, in order to enhance understanding of molecular mechanism between the two diseases, the semantic types of intermediate concepts mainly referred to the "Gene or Gene Product".

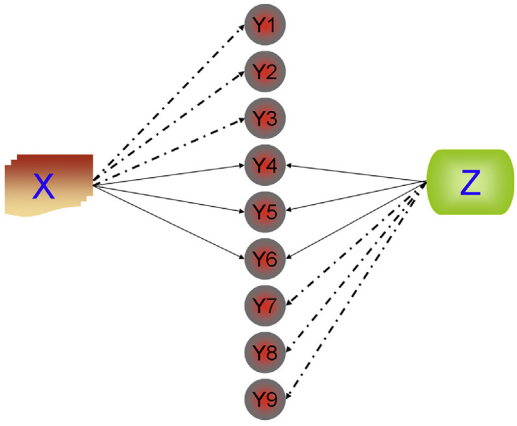

Fig. 1. Closed discovery process in the BITOLA system [16].
The search process starts simultaneously from X (eg.OLP) and Z (depression), resulting in overlapping Ys (potential mechanisms: play key roles in the communication between OLP and depression). The solid arrows indicate potentially interesting "gene or gene product" of discovery, and the dashed arrows indicate unsuccessful "gene or gene product".

2.2. Filtering of the intermediate concepts by gene expression profiles data

Because we mainly focused on "Gene or Gene Product" for the intermediate concepts, we tentatively filtered and evaluated the "Gene or Gene Product" through the overview of "Gene or Gene Product" expression levels (mRNA, Messenger Ribonucleic Acid) under different conditions (disease vs. control).

2.2.1. Gene expression profiles

Due to many complicated reasons, to date, there was no record of the gene expression profile data related to patient who had the clinical symptoms both in OLP and depression. We only detected whether the expression of "Gene or Gene Product" were different when disease status was compared to health status (OLP vs. control or depression vs. control), if yes, "Gene or Gene Product" would be kept for next analysis, and if not, it would be ruled out.

Gene expression data sets were obtained from Gene Expression Omnibus (GEO) database [20]. For OLP, the GSE52130 was developed from epitheliums (oral and genital lichen planus epithelium and normal oral and genital epithelium) of 13 patients diagnosed with oral and/or genital LP, and normal controls [21]. Since we only concentrated on the OLP, the samples from genital lichen planus epithelium were excluded. The other data GSE38616 was developed from oral mucosa of patients with OLP and healthy individuals [22]. As to depression, GSE54562, GSE54563, GSE54564, GSE54565, GSE54566, GSE54567, GSE54568, GSE54570, GSE54571, GSE54572 and GSE54575 were used [23]. More detailed information of these data can be seen in Table 1.

2.2.2. Statistical analysis of microarray data

The differentially expressed genes (DEGs) between diseases' status and the normal controls were analyzed using the web tool GEO2R [24]. Samples were assigned to the case or control group depending upon the sample source and experimental classification, respectively. T-test was used to sort out the DEGs. The top 250 probe was selected for following analysis, and finally probes were converted into genes' name.