Biochemical Pathway and Protein-Network Analysis of Dental Caries Based on Systems Biology Approaches

Zeinab Mohamadi¹, Zahra Khamverdi¹, Amir Taherkhani²

¹Department of Operative Dentistry, Dental School, Dental Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
²Research Center for Molecular Medicine, Hamadan University of Medical Sciences, Hamadan, Iran

Abstract

Background: Tooth decay (TD) is a multifactorial disorder, and several factors are involved in its etiology.

Objective: The present study aimed to unravel the main genes and molecular mechanisms underlying TD.

Methods: The dataset GSE1629 in the Gene Expression Omnibus (GEO) database was analyzed to uncover differentially expressed genes (DEGs) in patients with TD compared to patients with sound teeth. A protein-protein interaction network was built, and the most important clusters, hubs, transcription factors (TFs), and protein kinases involved in the regulation of TFs were disclosed. Signaling pathways and Gene Ontology terms dysregulated in TD were also identified.

Results: A total of 196 DEGs were determined (false discovery rate < 0.001; |Log2 fold change| > 1). PTPRC, ITGB2, TYROBP, MMP9, CXCL8, CD44, CCL2, C1QB, C3, and SPP1 were considered hub genes. Further, BPTF and MAPK1 were demonstrated to be the highest TFs and protein kinases likely involved in the pathogenesis of TD, respectively.

Conclusion: PTPRC, ITGB2, TYROBP, MMP9, CXCL8, CD44, CCL2, C1QB, C3, SPP1, BPTF, and MAPK1 may be regarded as potential markers for the therapeutic purposes of TD.

Keywords: Biomarker, Dental caries, Gene regulatory network, Protein-protein interaction network, Tooth decay

Background

According to a report about tooth decay (TD), the Global Burden of Disease study estimated that more than two billion of the world’s population are suffering from this highly prevalent chronic disorder (1). TD has remained a dominant global health issue in most modern and developing countries, in which 60%-90% of schoolchildren and a considerable number of adults have already been affected by this disease (2,3), TD is a comprehensive disorder, and various factors are involved in its occurrence. The pathogenesis of the disease can be elucidated by a logic diagram, which includes three circles and the interactions between them. The circles represent microbial load (plaque), intake of food, as well as host factors (e.g., genetic factors). The interaction among these circles could lead to dental caries (DCs). The parameter of “time” has also been added to the above circles, which displays the duration of interaction between caries-associated factors. Previous studies have shown that Streptococcus mutans and lactobacilli are primarily involved in the beginning and development of DC, respectively. Fermentable sugars are the primary substrates for these bacteria, and the carbohydrates produced from bacteria are mostly stable in the biofilm. DC could cause other oral disorders such as tooth pain, tooth defects, loss of teeth, and tooth crowding (4-12). However, the dentine–pulp complex has been reported to have regenerative characteristics by secreting a tertiary dentine extracellular matrix following a tooth injury, protecting the cells beneath the lesion, and maintaining the vitality of the tissues (13). Since previous studies have documented that the inflammatory reaction occurs due to bacterial infection in DC (14-21), further investigation is inevitable to illustrate the exact molecular mediation.

Reanalyzing the microarray datasets containing enormous gene expression data provides an opportunity to discover new genes previously not linked with the disorder and, therefore, could identify new potential biomarkers for diagnosis and curative aims. In addition, identifying novel caries-associated genes could help understand the etiology of this disorder (22,23). It may be hypothesized that the significant alteration in the expression of various genes in patients with DC compared to the individuals with sound teeth results in the deregulation of several

© 2023 The Author(s); Published by Hamadan University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
signaling pathways and biological processes (BPs) associated with the pathogenesis of DC. Moreover, the most critical genes and their regulating markers involved in the etiology of DC could be identified by analyzing the protein-protein interaction (PPI) network and gene regulatory network (GRN) associated with the disease, respectively (24). Kinase enrichment analysis also provides an opportunity to illustrate protein kinases involved in the phosphorylation of transcription factors (TFs) regulating the critical genes (25).

The present study, therefore, aimed to determine (1) differentially expressed genes (DEGs) within the pulpal tissue obtained from healthy and severely carious human teeth, (2) the hub genes in the PPI network, (3) the most significant TFs regulating the hub genes, (4) protein kinases involved in the TFs activity, and (5) signaling pathways and gene ontology (GO) in patients with tooth caries compared to individuals with sound teeth. This was performed by reanalyzing the gene expression dataset GSE1629 (22).

The gene dataset GSE1629, containing almost 15000 human sequences, was established by McLachlan et al (22) to identify genes associated with carious lesions to promote diagnostic and therapeutic approaches. McLachlan et al (22) performed their study by utilizing pooled RNA extracted from patients with DC and individuals with sound teeth extracted from the pulpal region. The completely decayed and healthy premolar and molar teeth, collected for orthodontic aims from individuals from age 20 to 30 years, were extracted from the Birmingham Dental Hospital. DC illustrated the carious lesions varying from enamel to decays. Next, the extracted teeth were instantly overwhelmed in RNA stabilizing buffer using the RNA Later (Sigma, UK), and the complete tissue of the pulp was cautiously eliminated by using a sterile dental probe and forceps (26).

Materials and Methods

Microarray Expression Dataset Analysis

The dataset GSE1629 (2) was chosen for analyses from the Gene Expression Omnibus (GEO) (NCBI GEO, http://www.ncbi.nlm.nih.gov/geo) (27). The GSE1629 consisted of four pooled pulpal tissue samples from 11 patients with clinically DC showing deep dental lesions/pulp exposure (n=2) and 12 with sound teeth (n=2). This dataset was developed on the platform of Affymetrix Human Genome U133A Array (GPL96). DEGs between pulpal tissue of patients with DC compared to individuals with sound teeth were identified using the GEO2R tool (28) with the benchmark of |log2 fold change| > 1 as well as the false discovery rate (FDR) < 0.001. Then, the Benjamini and Hochberg procedure was applied to calculate corrected P values.

The Protein-Protein Interaction Network and Clustering Analyses

The Search Tool for the Retrieval Of Interacting Genes (STRING) knowledge database (version 11.0 STRING, http://string-db.org) (29) was utilized for identifying possible connections between DEGs. Next, unconnected genes were removed from the network. The network analyzer tool within the Cytoscape 3.9.0 (http://www.cytoscape.org/) (30) was used for calculating the centralities (e.g., degree and betweenness) of nodes, while clustering analysis was carried out using the molecular complex detection (MCODE) plugin. The benchmarks considered for each module to be statistically significant included the number of nodes > 10, Degree cutoff = 2, MCODE score > 3, Max depth = 100, and the k-score = 2. MCODE is frequently used for determining condensed zones in PPI networks named clusters (modules) (31). It has been demonstrated that clusters include genes that take part in common pathways or BPs. Moreover, most of the clusters contain a seed node known to be a vertex of the modules based on its existing biological role in living systems (32-34).

Functional Analysis

The GO analysis is frequently used for mining particular biological properties from a set of high-throughput data, including genome, transcriptome, and proteome (35). The Reactome database (https://reactome.org/) provides perceptive bioinformatics tools for illustrating and analyzing the pathways to support theoretical and experimental study and genomics data mining, which has been developed at the Ontario Institute for Cancer Research, Cold Spring Harbor Laboratory, the New York University of Medicine, and The European Bioinformatics Institute (36). In addition, Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/) is an online bioinformatics database providing an entire set of functional annotation tools for researchers to realize the biological concepts of multiple proteins and genes (37,38). DAVID was used to illustrate the DEGs enrichment of cellular components (CCs) and molecular functions (MFs) as well as the main clusters enrichment of BPs. In addition, the Reactome database was used to demonstrate pathways significantly associated with the salient modules in the PPI network. The cutoff conditions for GO annotations and pathways were set to FDR < 0.05 and the enriched gene numbers ≥ 2.

Master Regulators of the Hub Genes

The iRegulon plugin was utilized for the prediction of upstream TFs linked to the hubs. Only TFs with a normalized enrichment score (NES) > 3.0 were considered to be statistically significant (39). The NES shows the significance of the detected TFs and a correlation with the area under the curve value associated with the motifs (40-42). Subsequently, significant motifs (e.g., GRN) consisting of TFs and their downstream hub genes were built using the iRegulon plugin. A GRN provides valuable information associated with the upstream regulators of critical genes involved in the pathogenesis of the disease.
Therefore, GRN analysis might lead to identifying several crucial genes regulated with common TFs. Thus, future studies could recognize TFs as novel drug targets (43).

**Consensus Sequences of Transcription Factors**

JASPAR is an open-access database ([https://jaspar.genereg.net/](https://jaspar.genereg.net/)) (44) providing helpful information about TFs binding sites across six taxonomic classes, including fungi, insect, nematode, Plantae, Urochordata, and Vertebrata. This useful web server includes consensus sequences Logos and position frequency matrices of several TFs. Position frequency matrices illustrate the occurrences of nucleotides at each position in a set of demonstrated TF-DNA interactions and could be used for constructing position-specific scoring matrices. The consensus sequences of these matrices are made of nucleotides with the highest frequency at each position. The consensus sequences matching scores for top-ranked TFs were calculated in R programming (version 4.0.0) (45) based on the method described by Xiong (46). The probability of consensus sequences matching the binding sites of TFs is described as \(2^{\text{match score}}\) times more than that by random chance (46).

**Protein Kinases Enrichment Analysis**

Protein kinases add a phosphate group from adenosine triphosphate to three residues of proteins, including Thr, Ser, and Tyr, while phosphatases catalyze the reverse reaction. Phosphorylation and dephosphorylation of proteins affect the activity, localization, stability, and interaction of substrates with other molecules in biological networks. The abnormal expression or activity of these two classes of enzymes could affect several signaling pathways and BPs in human cells, resulting in many disorders (47-49). Here, the online kinase enrichment analysis 3 (KEA3) knowledge database ([https://maayanlab.cloud/kea3/](https://maayanlab.cloud/kea3/)) (50) was utilized for uncovering upstream protein kinases potentially involved in the phosphorylation of TFs involved in the hub gene regulation. The KEA3 uses several types of resources to illustrate upstream TFs for a set of genes (proteins). These resources are based on PPI, kinase-substrate interaction, co-occurrence, and transcript co-expression (50). In this regard, the PPI libraries are as follows: BioGRID (51), mentha (52), huMAP (53), prePPI (54), MINT (55), HIPPIE (56), PPs (57,58), PSOPIA (59), REACTOME (60), ChengPPI (61), and STRING (62). In addition, PhosphositePlus (63), PhosD (64), PhosphoNetworks (65), PTMsigDB (66), ChengPPI (61), and Phospho.ELM (67) are libraries used to study kinase-substrate interactions. After a gene set is uploaded into the KEA3, upstream protein kinases are scored and ranked based on different algorithms used in different libraries. KEA3 then calculates the mean and the sum of the ranks in different libraries.

**Results**

**The Identification of Differentially Expressed Genes in Patients With Dental Caries**

In our present study, we obtained the data of four pooled samples from the pulp tissue of 11 patients with DC showing deep dentinal lesions/pulp exposure \((n=2)\) and 12 individuals with sound teeth \((n=2)\). A total of 196 DEGs, including 146 overexpressed and 50 underexpressed genes, with an FDR less than 0.001 and the \([\text{Log2 FC}] > 1\) were identified using the GEO2R online tool and considered for further analysis in the present study. Top-ranked over- and underexpressed genes in carious teeth compared to those in the sound teeth are presented in Figure 1, and Figure 2a demonstrates the volcano plot of DEGs. Figure 2b illustrates the hierarchical clustering of the hub genes, and Figure 2c demonstrates the interactions between hub genes obtained from the DAVID database.

**Protein-Protein Interaction Network, Clustering, and Functional Analyses**

A PPI network was achieved based on the STRING database’s DEGs, and the cut-off for confidence score was set to \(\geq 0.4\). After removing disconnected nodes, a PPI network with 141 genes and 917 connections was transferred into the Cytoscape software for advanced analyses.

The MCODE plugin identified two substantial clusters \((i.e., \text{cluster No. 1 and cluster No. 2})\) within the PPI network (Figure 3). These clusters were involved in the pathways and BPs associated with DC. They consisted of several genes mainly enriched in pathways and BPs linked to the human immune system, including ‘antigen processing and presentation of peptide or polysaccharide antigen via major histocompatibility complex (MHC) class II (BP), ‘immune response (BP),’ ‘antigen processing and presentation of exogenous peptide antigen via MHC class II (BP),’ ‘antigen processing and presentation (BP),’ ‘inflammatory response (BP),’ ‘leukocyte migration (BP),’ ‘interferon-gamma-mediated signaling pathway (BP),’ ‘T cell costimulation (BP),’ ‘chemotaxis (BP),’ ‘positive regulation of T cell activation (BP),’ ‘MHC class II antigen presentation (Pathway),’ ‘T cell receptor (TCR) signaling (Pathway),’ and ‘interferon-gamma signaling (Pathway).’

At an FDR less than 0.05, a total of 35 pathways and 14 BPs were found to be significantly affected in DC. Moreover, DAVID analysis revealed that a total of 26 CCs and 10 MFs were significantly dysregulated in patients with DC. Figure 4 illustrates the top 10 significant pathways, BPs, CCs, and MFs enriched in DC. Accordingly, the most meaningful BPs and pathways were linked to the human immune system, including ‘antigen processing and presentation of peptide or polysaccharide antigen via MHC class II’ and ‘MHC class II antigen presentation (Pathway).’ Furthermore, the average centrality value within the PPI network for betweenness and degree
was calculated as 0.012620178 and 13.0071, respectively. Accordingly, a total of 10 genes were revealed to have a degree and betweenness values more than twice of the nodes within the network, and therefore, were considered hub genes associated with the pathogenesis of TD (Table 1).

**Master Regulators of Hub Genes**

The iRegulon plugin was executed to predict TFs regulating the hub genes. Only TFs with the criteria of NES > 3.0 were considered to be statistically meaningful (39). Accordingly, there were 50 TFs significantly associated with tooth caries. The most enriched TF was bromodomain and PHD finger-containing transcription factor (BPTF) with an NES equal to 7.167. A total of four hubs were found to be downstream genes of BPTF including ITGB2, C1QB, CD44, PTPRC, and C1QB. The
Pathway and network analysis of dental caries

Figure 3. Module Analysis. Note. PPI: Protein-protein interaction; DEG: Differentially expressed genes; MCODE: Molecular complex detection. The network was built based on the DEGs in patients with dental caries compared to patients with sound teeth. The MCODE plugin discovered two salient clusters in the PPI network. The Hexagons illustrate seed nodes in each module. The size of the nodes is directly correlated with the clustering coefficient value of the genes in the leading PPI network.

Figure 4. Top-10 (a) Pathways, (b) Biological Processes, (c) Cellular components, and (d) Molecular Functions Enriched in Patients with DCs Concerning Their FDR. Note. DCs: Dental caries; FDR: False discovery rate. The genes’ names are presented at the top left corner of the Logos. The consensus sequences scores are also demonstrated below the names. The x-axis shows the name of the term, and the y-axis corresponds to the Log10 FDR.
statistics of the significant TFs are presented in Table 2.

**Binding Sites Logos and Matching Scores of Transcription Factors**

The top-ranked TFs with the criteria of NES > 5 were investigated in the JASPAR webserver. The consensus sequences Logos of nine TFs, including TCF12, SPI1, ESR2, STST3, RELB, PAX5, RXRA, TCF15, and HNF4A, were available in the database. The maximum and minimum scores for TFs binding sites were calculated at 22.91 and 7.5 for ESR2 and SPI1, respectively (Figure 5).

**Upstream Protein Kinases**

Top-10 protein kinases based on the mean of the ranks in different libraries were considered significant. Moreover, all top-10 ranked kinases had a P value less than 0.05. Mitogen-activated protein kinase 1 (MAPK1) demonstrated the best result with the mean rank score and a P value of 19.64 and 6.96E-05, respectively (Figure 6 and Table 3). Finally, a GRN was constructed based on the eight hub genes, 50 TFs, and 10 protein kinases. This regulatory network included 440 edges (Figure 7).

**Discussion**

DC is a common chronic disease affecting more than 60% of childhood and a considerable percentage of adults within the most industrialized and developing countries. It is a multifactorial disorder, and several agents are involved in its pathogenesis, including cariogenic bacteria and genetic factors (4-9). Despite many types of study in this scope, our knowledge regarding the molecular mechanisms and most essential genes that participate in the beginning and development of TD is limited (22).

In the present study, we tried to determine potential biomarkers, master regulators, protein kinases regulating TFs, signaling pathways, and BPs associated with the pathogenesis of DC. Therefore, gene expression data using pooled RNA from the pulp tissue of 12 individuals with

---

**Table 1.** A Total of 10 Hub Genes With the Criteria of Degree and Betweenness More Than Twice the Average of the Nodes in the PPI Network Associated With DCs

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Degree</th>
<th>Betweenness</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTPRC</td>
<td>53</td>
<td>0.0924642</td>
</tr>
<tr>
<td>ITGB2</td>
<td>49</td>
<td>0.0741381</td>
</tr>
<tr>
<td>TYROBP</td>
<td>42</td>
<td>0.0484431</td>
</tr>
<tr>
<td>MMP-9</td>
<td>40</td>
<td>0.0777514</td>
</tr>
<tr>
<td>CCL2</td>
<td>34</td>
<td>0.0660737</td>
</tr>
<tr>
<td>C1QB</td>
<td>34</td>
<td>0.0258517</td>
</tr>
<tr>
<td>C3</td>
<td>31</td>
<td>0.0308056</td>
</tr>
<tr>
<td>SPP1</td>
<td>30</td>
<td>0.06125618</td>
</tr>
</tbody>
</table>

Note: ID: Identification; DCs: Dental caries; PPI: Protein-protein interaction; ITGB2: Integrin beta 2; MMP: Matrix metalloproteinases.

**Table 2.** The 50 TFs Significantly Identified for Regulating Hub Genes in DCs

<table>
<thead>
<tr>
<th>TF</th>
<th>Targets</th>
<th>NES</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPTF</td>
<td>ITGB2, CD44, PTPRC, C1QB</td>
<td>7.167</td>
</tr>
<tr>
<td>TROVE2</td>
<td>C1QB, C3, CD44, SPP1, PTPRC, CCL2</td>
<td>7.113</td>
</tr>
<tr>
<td>TCF12</td>
<td>C1QB, C3, ITGB2, CD44</td>
<td>6.472</td>
</tr>
<tr>
<td>SPI1</td>
<td>C1QB, PTPRC, C3, CD44, ITGB2, TYROBP</td>
<td>6.027</td>
</tr>
<tr>
<td>CBF1</td>
<td>SPP1, ITGB2, TYROBP</td>
<td>5.878</td>
</tr>
<tr>
<td>ESR2</td>
<td>ITGB2, C3, C1QB, SPP1, CD44</td>
<td>5.419</td>
</tr>
<tr>
<td>STAT3</td>
<td>PTPRC, C3</td>
<td>5.402</td>
</tr>
<tr>
<td>RELB</td>
<td>C3, CD44, ITGB2, C1QB, CCL2, PTPRC</td>
<td>5.355</td>
</tr>
<tr>
<td>PAX5</td>
<td>PTPRC, ITGB2</td>
<td>5.319</td>
</tr>
<tr>
<td>RXRA</td>
<td>CD44, ITGB2, C3, CCL2, TYROBP, C1QB</td>
<td>5.155</td>
</tr>
<tr>
<td>TCF15</td>
<td>SPP1, CD44, ITGB2, PTPRC, C1QB</td>
<td>5.133</td>
</tr>
<tr>
<td>ZNF691</td>
<td>CCL2, C3, CD44, ITGB2, PTPRC, C1QB, TYROBP</td>
<td>5.124</td>
</tr>
<tr>
<td>HNF4A</td>
<td>ITGB2, C3</td>
<td>5.008</td>
</tr>
<tr>
<td>PREB1</td>
<td>C1QB, ITGB2</td>
<td>4.97</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>CCL2, C1QB</td>
<td>4.881</td>
</tr>
<tr>
<td>FOXO3</td>
<td>PTPRC, ITGB2, CCL2, SPP1, CD44</td>
<td>4.861</td>
</tr>
<tr>
<td>EP300</td>
<td>ITGB2, C1QB, PTPRC, CD44, CCL2</td>
<td>4.506</td>
</tr>
<tr>
<td>TFAP2C</td>
<td>SPP1, C3, ITGB2</td>
<td>4.478</td>
</tr>
<tr>
<td>SPI8</td>
<td>CD44, CCL2, C3, C1QB, TYROBP, ITGB2, PTPRC</td>
<td>4.461</td>
</tr>
<tr>
<td>JAZF1</td>
<td>C1QB, ITGB2, CD44</td>
<td>4.429</td>
</tr>
<tr>
<td>TFAP2A</td>
<td>ITGB2, SPP1, C3</td>
<td>4.421</td>
</tr>
<tr>
<td>HSF1</td>
<td>CD44, PTPRC, CCL2, C1QB, C3, ITGB2</td>
<td>4.402</td>
</tr>
<tr>
<td>PX8</td>
<td>TYROBP, ITGB2, C3</td>
<td>4.347</td>
</tr>
<tr>
<td>GATA1</td>
<td>C1QB, ITGB2, PTPRC</td>
<td>4.343</td>
</tr>
<tr>
<td>PROX2</td>
<td>ITGB2, CCL2, CD44, C1QB, SPP1, TYROBP</td>
<td>4.338</td>
</tr>
<tr>
<td>MEF2A</td>
<td>ITGB2, CD44</td>
<td>4.311</td>
</tr>
<tr>
<td>TBL1XR1</td>
<td>ITGB2, CD44</td>
<td>4.307</td>
</tr>
<tr>
<td>TEAD4</td>
<td>ITGB2, CCL2</td>
<td>4.198</td>
</tr>
<tr>
<td>GTF1B</td>
<td>C3, CCL2</td>
<td>4.184</td>
</tr>
<tr>
<td>POLE3</td>
<td>SPP1, C1QB, ITGB2, CD44, PTPRC</td>
<td>4.179</td>
</tr>
<tr>
<td>POLR3A</td>
<td>CD44, C1QB</td>
<td>4.099</td>
</tr>
<tr>
<td>TRMT1</td>
<td>C1QB, SPP1, ITGB2, C3, CD44</td>
<td>4.08</td>
</tr>
<tr>
<td>HNF4G</td>
<td>ITGB2, CCL2, C3, C1QB</td>
<td>4.061</td>
</tr>
<tr>
<td>DBP</td>
<td>CCL2, SPP1</td>
<td>4.011</td>
</tr>
<tr>
<td>SP1</td>
<td>C3, ITGB2</td>
<td>4.004</td>
</tr>
<tr>
<td>NFKB1</td>
<td>CD44, ITGB2</td>
<td>3.966</td>
</tr>
<tr>
<td>HOXB8</td>
<td>C3, ITGB2, PTPRC, SPP1, C1QB</td>
<td>3.821</td>
</tr>
<tr>
<td>STAT1</td>
<td>CCL2, C3</td>
<td>3.801</td>
</tr>
<tr>
<td>CELF5</td>
<td>C1QB, ITGB2</td>
<td>3.684</td>
</tr>
<tr>
<td>NFIC</td>
<td>CD44, CCL2, ITGB2</td>
<td>3.677</td>
</tr>
<tr>
<td>TAFP2B</td>
<td>PTPRC, ITGB2, SPP1, C1QB</td>
<td>3.644</td>
</tr>
<tr>
<td>HNRNPH1</td>
<td>C1, CCL2, C1QB</td>
<td>3.621</td>
</tr>
<tr>
<td>SMAD5</td>
<td>ITGB2, SPP1</td>
<td>3.566</td>
</tr>
<tr>
<td>GTF3C2</td>
<td>ITGB2, C1QB, C3</td>
<td>3.462</td>
</tr>
<tr>
<td>AVEY</td>
<td>PTPRC, CD44, SPP1, CCL2, ITGB2</td>
<td>3.457</td>
</tr>
<tr>
<td>ARID3A</td>
<td>ITGB2, C1QB, PTPRC</td>
<td>3.348</td>
</tr>
<tr>
<td>TGF212Y</td>
<td>CCL2, PTPRC, ITGB2, C3, C1QB</td>
<td>3.221</td>
</tr>
<tr>
<td>GMC1</td>
<td>ITGB2, C1QB, C3, CCL2, SPP1</td>
<td>3.14</td>
</tr>
<tr>
<td>SMC3</td>
<td>TYROBP, CD44</td>
<td>3.133</td>
</tr>
<tr>
<td>RAB7A</td>
<td>C3, C1QB, ITGB2, TYROBP, TYROBP</td>
<td>3.09</td>
</tr>
</tbody>
</table>

Note: DCs: Dental caries; BPTF: PHD finger-containing transcription factor; TF: Transcription factor; NES: Normalized enrichment score.
Pathway and network analysis of dental caries

clinically sound teeth and 11 patients with DC exhibiting deep dentinal lesions/pulp exposure were obtained from the GEO web tool and analyzed. A total of 196 DEGs, including 146 overexpressed and 50 underexpressed genes, with the criteria of FDR < 0.001 and the value of $|\text{Log2 FC}| > 1$ were identified. Further, the most significant pathways and BPs associated with DC were identified using the Reactome and DAVID online databases, respectively. Moreover, a PPI network associated with DC was built and analyzed based on DEGs.

The MHC includes a set of genes within the jawed vertebrates taking part in the immune system (68). The MHC is positioned at chromosome-6 (6p21.3) in humans and contains more than 200 genes and many alleles (69,70) and encodes glycoproteins named human leukocyte antigens (HLAs) that take part in the body’s immune defense by the presentation of short peptides to T cells (71).

Various HLA alleles are involved in different human disorders and conditions such as autoimmune disease, inflammation, cancer, social behavior, as well as a shorter lifetime. However, the exact underlying mechanism has not been clearly understood. Overall, HLAs are classified into two main subclasses (i.e., HLA class I and HLA class...
II) based on their different structures, biological roles, and the expression on the surface of cells.

All nucleated cells contain different amounts of HLA class I proteins on their cell membrane. These proteins contribute to presenting peptides from degraded proteins or abnormal ones from hazardous viruses inside the cytoplasm to the TCR of the CD8+ T cells, contributing to activating the defense system. Furthermore, the HLA class II proteins are exhibited on the surface of B cells and the other antigen-presenting cells, known as active immune cells. These proteins are involved in presenting extracellular antigens to CD4+ T cells. Additionally, T cells produce HLA-II main isoforms, including HLA-DP, HLA-DR, and HLA-DQ, almost 3-5 days after activation. Furthermore, interferon-γ and interleukin (IL)-4 are produced just after a few hours following the activation of TCR (72-80).

TD occurs after the deterioration of the tooth’s organic and/or mineral texture. Matrix metalloproteinases (MMPs) take part in collagen loss, while bacteria contribute to damaging the inorganic compounds within the healthy tissues of the teeth (81). S. mutans is the most prevalent caries-associated bacteria, but Lactobacillus acidophilus and Lactobacillus casei have also been linked with dental plaque (82,83).

A significant correlation has been observed between caries resistance and a robust immune system. In this regard, caries-free individuals revealed higher serum antibody titers to streptococcal antigens than caries-susceptible individuals (84). Further, caries-free individuals were...
found to produce T cells with a salient ability to increase at the time of exposure to S. mutans antigens (85). Moreover, caries-resistant subjects can potentially generate more active T-helper-cells upon exposure to a minor amount of S. mutans antigens compared with patients exhibiting TD (86). Therefore, HLA-related genes from cluster 2 might enhance the immune system against S. mutans antigens and vice-versa.

It has also been reported that different HLA alleles may affect bacterium-induced TD. In this case, a significant DR-positive cell aggregation was observed within the pulp tissue of the patients with superficial caries (87). Moreover, the DR-positive cell aggregation was expanded to odontoblast cells in patients with deeper caries lesions (86).

Acton et al (88) performed a study to examine the possible correlation between MHC alleles at the HLA-DRB1, DQB1, and TNFa microsatellite loci and the number of oral bacteria associated with DC, in addition to the decayed, missing, and filled teeth index in 186 African-American females. The patients had an average age and median decayed, missing, and filled surface indexes of 20.8 years and 9 (range 0–68), respectively. Acton et al (88), reported a significant correlation between enhanced levels of S. mutans and the elevated expression of DRB1*4 and DRB1*3 (P-value = 0.005). Furthermore, the enhanced level of DRB1*8 was linked to the increased level of S. mutans (P = 0.04). A positive correlation was also observed between the levels of DRB1*1 and L. casei (P = 0.04). Moreover, the TNFa allele 103 and TNFa 117 were negatively and positively associated with the levels of L. acidophilus, respectively. Moreover, Acton et al (88) demonstrated a significant correlation between the production of HLA class II alleles, TNFa genetic, and the colonization of DC-associated oral bacteria, including S. mutans, L. casei, and L. acidophilus.

Based on the obtained results, a total of 25 nodes were involved in cluster 1 such as HLA-DRB1 and HLA-DRA. Further, cluster 2 consisted of 19 genes, including HLA-DPB1, HLA-DQA1, HLA-DMB, HLA-DQB1, HLA-DPA1, and HLA-DMA. Further analysis revealed that the HLA isotypes mentioned above were significantly over-expressed in the pulp tissue of patients with tooth caries compared to patients with sound teeth with the benchmark of FDR < 0.001 and the absolute value of Log2 FC > 1. Accordingly, it may be hypothesized that these genes contribute to the beginning and/or developing of TD. Otherwise, the enhanced expression of these genes may be due to the response of increased tooth caries. However, more experiments are necessary to approve our findings.

A total of 10 genes demonstrated degree and betweenness values higher than twice the average of the genes within the PPI network and, therefore, were found to be hub genes playing a significant role in the etiology of DC. The top-ranked genes based on the degrees were PTPRC (CD45), ITGB2, TYROBP, and MMP9, while PTPRC, CD44, MMP9, and ITGB2 revealed the highest betweenness centrality in the PPI network, respectively.

CD45 is a tyrosine-protein phosphatase antigen encoded by the PTPRC gene. It positively regulates T cells through binding to the dipeptidyl-peptidase 4 (DPP4) (89-91). Lacerda-Pinheiro et al (92) designed a study to understand the principles of the differentiation of pulp cells promoted by the bioactive molecules. The authors implanted the agarose beads (alone or covered with the products of the amelogenin gene [A + 4 and A-4]) in the mucosa of the cheeks in mice. They reported that agarose increased the recruitment of CD45+cells, leading to the enhanced recruitment of leukocytes from the vascular compartment, resulting in the enhanced re-differentiation of leukocytes into osteo-chondrogenic lineages. In addition, the coated beads with A+4 induced the production of osteo-chondrogenic markers, including RP59, SOX9, and bone sialoprotein. It is worth mentioning that RP59 is an antigen produced in the bone marrow and young osteoblasts (93,94). Moreover, SOX9 is a TF that participates in chondrocyte differentiation and skeletal development (95). Our results revealed that CD45 was significantly overexpressed in the pulp tissue of patients with DC compared to patients with sound teeth (FDR = 3.46E-05; Log2 FC = 2.8), which may be due to the response of increased TD.

MMPs are a family of zinc-dependent proteolytic enzymes taking part in the degradation of extracellular matrix proteins such as collagens in their native and denatured forms. MMPs are activated by decreasing the pH to 4.5, followed by neutralization. Previous studies have demonstrated that the dentin matrix mainly contains type I collagen, and MMPs have a specific role in DC pathogenesis (96). Several types of MMPs are involved in dentin matrix elimination, including MMP8 (collagenase-2), MMP2 (72-kD gelatinsase), MMP9 (92-kD gelatinsase), MMP13 (collagenase-3), and MMP20, leading to TD (81,97); therefore, the systemic application of MMP inhibitors may be helpful in the prevention of TD (96). Wang et al (98) studied the correlation between the saliva levels of MMP2/MMP9 and different stages of DC in childhood. The levels of MMP2 and MMP9 were measured by the enzyme-linked immunosorbent assay. Wang et al (98) reported that the saliva levels of MMP9 in severe and mild-caries groups were significantly higher compared with the caries-free group (P value < 0.05). However, no significant difference was observed between the severe caries group and the mild caries group (P value >0.05). Wang et al (98) concluded that the salivary levels of MMP9 might be correlated with TD in children. According to our results, MMP9 was significantly overexpressed in the pulp tissue of patients with TD compared to patients with sound teeth (FDR = 5.78E-05; Log2 FC = 2.21), suggesting that MMP9 is involved in the etiology of DC.

Integrin beta 2 (ITGB2) is a type of integrin chain and has been reported to be produced explicitly in leukocytes, promoting the connection between leukocytes and the
endothelium (99,100). Zhang et al (101) compared the expression level of ITGB2 in cancer-associated fibroblasts (CAFs) and normal fibroblasts in patients with oral squamous cell carcinoma (OSCC) by using the reverse transcription polymerase chain reaction and western blot analyses, demonstrating that the ITGB2 expression significantly increased in CAFs compared with the matched normal fibroblasts. The authors reported that the over-expression of ITGB2 caused the hyper-activation of glycolysis through PI3K/AKT/mTOR pathways, leading to enhanced lactate excretion in CAFs and more proliferation of OSCC cells. The hyperactivity of glycolysis was found and confirmed through bioinformatics approaches and gas chromatography/mass spectrometry analysis, respectively. According to the present results, a significant increase was found in the expression of ITGB2 in the pulp tissue of patients with DC compared with patients with sound teeth (P value = 1.6E-04; Log2 FC = 1.97). We hypothesize that some of the mechanisms that mediate DC may be similar to those in OSCC, leading to more production and excretion of lactate from the pulp tissue into the dentin region, resulting in a lower pH; as a consequence, causing enhanced dental decay, although this requires confirmation.

A total of 50 TFs were determined as master regulators of eight of the hub genes. Bromodomain and BPTF were the most significant enriched TFs with the criteria of NES = 7.167, regulating ITGB2, CD44, PTPRC, and C10QB. BPTF is the central member of the human nucleosome remodeling factor (102). The misregulation of BPTF has been demonstrated in many human cancers such as bladder cancer (103), hepatocellular carcinoma (104), glioma (105), and lung cancer (106). However, the etiological role of BPTF in DC needs more studies in the future.

Moreover, MAPK1 demonstrated the best score among protein kinases regulating TFs (mean rank = 19.64). MAPKs are intracellular protein kinases necessary for inflammatory bone loss by activating MMPs and inflammatory cytokines (107). In this regard, p38 MAPK is an upstream activator of TNF-a, IL-1β and -6, and prostaglandin E2 (108), leading to the overexpression of prostaglandins, MMPs, and receptor activator of nuclear factor kappa beta (109,110), resulting in osteoclastogenesis and bone loss procedures (107). Similar mechanisms might be involved in the aberrant expression of MAPK1 in DC, leading to tooth loss; however, confirmation is needed.

Our study had certain limitations. Only individuals from the Birmingham Dental Hospital were included in the GSE1629 dataset; therefore, the present results may not entirely justify the data obtained from the other nationalities. Further, only four pooled pulpal tissue samples from 11 patients with clinically DC having deep dentinal lesions/pulp exposure and 12 individuals with sound teeth were included in the GSE1629 dataset, and the sample size was small. Hence, a greater number of individuals with DC and sound teeth may promote the reliability of statistical approaches and probably illustrate more DEGs that are significantly linked to the pathogenesis of carious teeth. Moreover, the genes profiled in this study were achieved based on bioinformatics analyses; therefore, considerable in vitro experiments are necessary to approve our findings. In future studies, molecular experiments with large targeted cohorts are obligatory to support our data.

Conclusion
Overall, the present study successfully identified 196 genes differentially expressed in the pulpal tissue of patients with tooth caries compared to patients with sound teeth (adjusted P value < 0.001; |Log2 FC > 1|). Furthermore, a total of 10 hub genes including PTPRC, ITGB2, TYROBP, MMP9, CXCL8, CD44, CCL2, C10QB, C3, and SPP1 were determined as proteins of potential importance to the pathogenesis of DC in the PPI network. Further, BPTF and MAPK1 were found to be the most significant master regulator and protein kinase potentially involved in the etiology of DC, respectively. Moreover, substantial modules in the PPI network were mostly enriched in the pathways and BPs linked to the immune system. However, more studies and wet-lab experiments must be carried out in the future to confirm these findings and demonstrate their definite role in the etiology of DC.

Acknowledgments
The authors would like to appreciate the Deputy of Research and Technology, Dental Research Center, and Research Center for Molecular Medicine, Hamadan University of Medical Sciences, Hamadan, Iran, for their support.

Authors’ Contribution
Conceptualization: Amir Taherkhani, Zeinab Mohamadi, Zahra Khamverdi.
Data curation: Amir Taherkhani.
Formal analysis: Amir Taherkhani, Zeinab Mohamadi.
Investigation: Amir Taherkhani.
Methodology: Amir Taherkhani, Zeinab Mohamadi.
Project administration: Amir Taherkhani.
Resources: Amir Taherkhani, Zeinab Mohamadi, Zahra Khamverdi.
Supervision: Amir Taherkhani.
Validation: Amir Taherkhani.
Visualization: Amir Taherkhani.
Writing – original draft: Amir Taherkhani.
Writing – review & editing: Zahra Khamverdi, Zeinab Mohamadi.

Competing Interests
The authors declare that they have no competing interests.

Data Availability Statement
Pathway and network analysis of dental caries

Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/), and ‘Reactome online database (https://reactome.org’).

Ethical Approval
The present study has been confirmed by the Ethics Committee of Hamadan University of Medical Sciences, Hamadan, Iran (Ethics no. IR.UMSHA.REC.1399.493).

Funding
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References


Pathway and network analysis of dental caries


10.1099/14171-2164-7-322. doi: 10.1186/1743-8089-8-10


Zahid H, Olson NM, Pomerantz WCK. Opportunity knocks for


