Investigation of pathogenic genes in peri-implantitis from implant clustering failure patients: A whole-exome sequencing study

Soo Hyung Lee

The Graduate School
Yonsei University
Department of Dental Science
Investigation of pathogenic genes in peri-implantitis from implant clustering failure patients: A whole-exome sequencing study

(Directed by Prof. Dong-Hoo Han, D.D.S., M.S.D., Ph.D.)

A Dissertation Thesis
Submitted to the Department of Dental Science
And the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Dental Science

Soo Hyung Lee

June 2014
This certifies that the dissertation thesis of Soo Hyung Lee is approved.

Thesis supervisor: Dong-Hoo Han

Thesis committee: Jin Kim

Thesis committee: Jae Hoon Lee

Thesis committee: Jee Hwan Kim

Thesis committee: Ji Hyun Lee

The Graduate School
Yonsei University
June 2014
감사의 글

본 논문이 완성되기까지 오랜 시간 지도와 격려로 저를 이끌어주신 한동후 지도 교수님께 진심으로 감사를 드립니다. 논문 심사과정에서 세심한 지도와 가르침을 주신 김진 교수님, 이재훈 교수님, 김지환 교수님, 이지현 교수님께 깊은 감사를 드립니다. 논문 전반에 걸쳐 많은 도움을 주신 김지영 연구원에게도 감사의 마음을 전합니다. 지면을 통해서 일일이 연급하시는 못하지만 저에게 도움과 격려를 주신 모든 분들께 다시 한번 진심으로 감사 드립니다.

마지막으로 항상 저를 지켜봐 주시는 아버지, 어머니, 장인어른, 장모님께 감사 드리며, 제가 하는 일에 언제나 전념할 수 있도록 세심하게 배려해주신 아내 윤주와 하루가 다르게 성장해 나가면서 저에게 큰 기쁨이 되고 있는 딸 채민에게도 고마움과 사랑을 전하며 이 논문을 나누고자 합니다.

2014 년 06 월

이수형 드림
# Table of contents

List of Figures ......................................................... ii
List of Tables ......................................................... ii
Abstract ................................................................. 1
I. Introduction ......................................................... 4
II. Materials and Methods ........................................... 8  
   1. Ethic statement ................................................. 8  
   2. Patient selection .............................................. 8  
   3. Existing comparing data set ................................. 9  
   4. Sample collection ............................................ 10  
   5. Whole exome sequencing on HISEQ 2000 using SureSelect All Exon kit 50Mb .......... 10  
   6. Whole exome sequencing and variant analysis .......... 12  
   7. Statistical analysis ........................................... 15  
   8. Gene set enrichment analysis ............................... 15  
   9. Construction of protein functional network .............. 16  
III. Results .............................................................. 17  
   1. Whole exome sequencing and variant filtering findings ................................................. 17  
   2. Bioinformatics findings: GSEA and network analysis ....................................................... 19  
   3. Known peri-implantitis genes ................................ 20  
IV. Discussion .......................................................... 25  
V. Conclusion .......................................................... 31  
References .............................................................. 32  
Abstract (In Korean) ............................................... 37
List of Figures

Fig. 1. The filtration and prioritization framework used for data analysis...........1 4
Fig. 2. Analysis of gene function enrichment and construction of functional
network........................................................................................................2 3

List of Tables

Table Ⅰ. Patient Information................................................................................9
Table Ⅱ. Summary of number of reads and coverage ........................................1 8
Table Ⅲ. Number of genes common to the 6 affected individuals in each filtering
step....................................................................................................................1 8
Table Ⅳ. Top 10 clusters ranked by enrichment score from GSEA results............2 1
Table Ⅴ. Known implantitis genes found in all variants .....................................2 4
Abstract

Investigation of pathogenic genes in peri-implantitis from implant clustering failure patients:

A whole-exome sequencing study

Soo Hyung Lee D.D.S., M.S.D.

(Directed by Prof. Dong-Hoo Han, D.D.S., M.S.D., Ph.D.)

Purpose: Peri-implantitis is a frequently occurring gum disease linked to multi-factorial traits with various environmental and genetic causalities and no known concrete pathogenesis. The varying severity of peri-implantitis among patients with relatively similar lifestyle and environmental traits suggests a genetic aspect which needs to be investigated to understand and regulate the pathogenesis of the disease. The purpose of this study is to find associated pathogenic genes of peri-implantitis from implant clusterization failure patients via WES, GSEA, and network analysis.

Materials & Methods: Six unrelated individuals with multiple implant clustering failure due to severe peri-implantitis were chosen for this study. These six individuals had relatively healthy lifestyle with no environmental causalities affecting peri-implantitis.
With limitations in mind, genetic study approach was taken to investigate pathogenic genes in peri-implantitis. Whole-exome sequencing (Agilent SureSelect Human All Exon 50 Mb kits) was performed on collected saliva samples via OG-500 (DNAgenoTeK). Common variants with minor allele frequencies (MAF) $\geq 0.05$ from all control datasets were eliminated and the variants having high and moderate impact and loss of function were used for comparison. Gene set enrichment analysis was performed to reveal the functional groups associated with genetic variants. Network analysis was applied to find relationships between functional clusters.

**Results:** 2,022 genes were left after filtering against dbSNP, the 1000 Genomes East Asian population, and healthy Korean randomized subsample data (GSK project). 175 (p-value $< 0.05$) out of 927 gene sets were obtained via GSEA (DAVID). The top 10 was chosen (p-value $< 0.05$) from cluster enrichment showing significance of cytoskeleton, cell adhesion, and metal ion binding. Among the functional groups, metal ion binding was located in the center of all clusters, indicating dysfunction of regulation in metal ion concentration might affect cell morphology or cell adhesion, resulting in implant failure.

**Conclusion:** Our results suggest that various genes and gene sets related to factors involved in cell adhesion such as cadherin, fibronectin, EGF domains, and cytoskeletons play critical roles in the osseointegration and pathogenesis of peri-implantitis. Interestingly, these two gene sets are indirectly linked via the metal ion binding protein. One may conclude that dysfunction in cell morphology and cell adhesion by regulatory
imbalance in metal ion concentration and dysregulation of integrins which can have impact on surface adhesion would affect the occurrence of peri-implantitis.

Key words: Peri-implantitis, Pathogenic genes, Implant clustering failure, Whole-exome sequencing, Gene set enrichment analysis, Network analysis, Cell adhesion, Cytoskeleton, Metal ion concentration
Investigation of pathogenic genes in peri-implantitis from implant clustering failure patients: A whole-exome sequencing study

Soo Hyung Lee D.D.S., M.S.D.

The Graduate School
Yonsei University
Department of Dental Science

(Directed by Prof. Dong-Hoo Han, D.D.S., M.S.D., Ph.D.)

I. Introduction

The successful incorporation of titanium implants in the rehabilitation of edentulous patients has been well-investigated over the years. According to statistics, over 2 million dental implants are installed annually, and that could rise over the next few years. Despite the high success rates of osseointegrated implants, biological complications related to implants are growing in numbers and these complications often result in implant loss.
Failure rate of dental implant that is currently reported within first 10 years are approximately 3-5%\textsuperscript{5,6}. Peri-implantitis is a biological complication that occurs in dental implant patients and comprises a range of destructive inflammatory processes affecting the soft and hard tissues surrounding dental implants for which there is no current gold-standard treatment\textsuperscript{7}.

Diagnosis of peri-implantitis is based on changes of color in the gingival, bleeding and probing depth of peri-implant pockets, suppuration, X-ray, and gradual loss of bone height around the implant\textsuperscript{8}. Despite all these diagnostic methods, we need to note that peri-implantitis can develop without any obvious symptoms such as pain\textsuperscript{9,10} so that patients often fail to notice the development of disease. Due to difficulties in early detection of peri-implantitis, implant failures are steadily increasing with clinical studies reporting near 4% implant loss\textsuperscript{11-13}. Hence we need a brand-new approach for diagnosing peri-implantitis.

Previous studies on peri-implantitis have found that peri-implantitis and implant failures cluster in subsets of individuals, and that a patient who has lost one implant is at elevated risk of experiencing other implant losses\textsuperscript{14,15}. Such phenomenon is called implant clustering failure which is the failure of more than one implant in a patient, not necessarily in the same area or quadrant. In a previous literature review, findings suggested that implant failures are not randomly distributed in the treated populations, but rather cluster in specific high risk groups and individuals\textsuperscript{16}. A few individuals have a concentrated risk for multiple implant failure and subsequently experience multiple losses. There are limitations in peri-implantitis pathogenesis research due to multiple casualties affecting the disease. The effects of both multiple genes and environmental factors would
play a critical role in peri-implantitis. With careful consideration of study population such as choosing case group with distinctive phenotype of the disease, relatively healthy lifestyle and similar environmental traits, patients with implant clustering failure was considered for current study. Among individuals sharing relatively similar environmental factors, implant clustering failure has low occurrence among total population, suggesting its strong association to genetic factors rather than environmental ones. The cluster phenomenon supports evidence that specific host characteristics, such as genetic factors, play an important role in bone resorption and in the development of peri-implantitis leading to implant failure 17.

Therefore, early diagnosis of peri-implantitis through the detection of pathogenic genes in advance of visual symptoms and radiographic findings would prevent implant failure due to severe peri-implantitis and increase the implant success rate as well. So we tried to find associated pathogenic genes of peri-implantitis from implant clustering failure patients in this study.

There have been several studies on the role of cytokines in peri-implantitis 18. A recent systematic review of the association between genetic predisposition and biological complications of dental implants suggested that there is no strong association among specific genetic polymorphisms (IL-1A, IL-1B, IL-17RC) and implant failure, although there was a notable tendency indicating a link between the IL-1 genotype and peri-implantitis 19,20. Other previous researches have presented MMP-8 and PGE-2 which are regulated by IL-1 as possible genetic markers for unsuccessful implants based on their role in regulation of extracellular matrix (ECM) which can possibly enhance bone healing within defects and promote implant osseointegration 21. Nonetheless, previous researches
have not elucidated genetic association of peri-implantitis suggesting new approach in genetic association study.

Whole Exome Sequencing (WES) in genetics had the greatest impact on Mendelian disorders, distinguishing more than 100 genes in rare Mendelian disorders between 2010 and 2012. Approaches in investigating novel genetic mechanisms, phenotypic variability, modifier genes, allelic variants, and genetic variations in Mendelian disorders may also elucidate complex disorders such as peri-implantitis. Unlike Mendelian disorders, peri-implantitis is a multi-factorial disorder comprising complex disorders, multiple genes, as well as lifestyle or environmental factors. To maximize the specificity of our results, we selected patients with clusterization failure due to peri-implantitis. WES alone may not provide pragmatic results on the association between a specific disease and genetic variants due to the extensive raw data, pointing out the need to incorporate bioinformatic data management, computational analyses, and data mining. Incorporation of Gene Set Enrichment Analysis (GSEA) and protein functional network study on WES data may determine genetic variants explicitly related to that disease.

The objective of this study is to find associated pathogenic genes of peri-implantitis from implant clusterization failure patients via WES, GSEA, and network analysis.
II. Materials and Methods

1. Ethic Statement

All research involving human subjects or human data was approved by the Institutional Review Board of Yonsei University College of Dentistry (Yonsei IRB No. 2-2012-0023). All clinical investigation was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants before enrolling in this study.

2. Patient Selection

Six individuals with implant clustering failure due to severe peri-implantitis aged between 50 and 68 were analyzed by massively parallel sequencing in this study (1 male, 5 female).

Among the patients had been treated with osseointegrated titanium implants for partial or complete edentulism in the Implant Clinic of Yonsei University Dental Hospital by one prosthodontist from February 1991 to October 2013, ninety seven patients had been experienced implant failure and the number of failed implants was 152. The chosen six patients were included as the part of the ninety seven patients and showed uniquely implant clustering failure. These patients had no history of para-funtional habits, smoking experience, periodontal disease, or systemic disease such as diabetes mellitus and osteoporosis. The number of inserted implants to the patients was 55, 26 of which were
explanted. Presence of mobility, vertical resorption of less than 50%, and pus discharge on remained implants yielded a poor prognosis in all selected cases (Table I).

**Table I.** Patient Information

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Number of implants placed</th>
<th>Number of implant failures</th>
<th>History of systemic diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>67</td>
<td>14</td>
<td>7</td>
<td>Penicillin allergy</td>
</tr>
<tr>
<td>Female</td>
<td>68</td>
<td>11</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Female</td>
<td>53</td>
<td>6</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Male</td>
<td>54</td>
<td>6</td>
<td>2</td>
<td>Hepatitis</td>
</tr>
<tr>
<td>Female</td>
<td>63</td>
<td>6</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Female</td>
<td>50</td>
<td>12</td>
<td>7</td>
<td>None</td>
</tr>
</tbody>
</table>

3. **Existing Comparing Data Set**

126 Koreans from the GSK project aged between 23 and 46 (109 male, 17 female) were chosen as randomized subsample from reference population along with dbSNP137 and the 1000 Genomes East Asian population. Exome data of Korean randomized subsample, dbSNP137, and 1000 Genomes East Asian population were used to eliminate
common variants from the six selected patients’ WES data. The 126 Korean randomized subsample, healthy Koreans regardless of gender and age, had originally been recruited for a thyroid cancer study (GSK project). This group had no history of diseases known to affect periodontal disease, such as diabetes and osteoporosis.

4. Sample Collection

To obtain DNA samples, each patient’s saliva was collected using OG-500 (DNA Genotek, Ottawa, Ontario, Canada, Cat. #OG-500). OG-500 kit instructions were followed: first, all 6 individuals were asked to collect 2 mL of saliva in the tube of an Oragene DNA Self-Collection kit containing 2 mL of DNA-preserving solution. The lid was closed to release the storing liquid to mix with the saliva. Genomic DNA collection, DNA extraction, and further analysis were performed by DNA Link Inc., Seoul, South Korea.

5. Whole Exome Sequencing on HISEQ 2000 using SureSelect All Exon kit 50Mb

With an OD260/280 ratio of 1.8-22, DNA should be as intact as possible. Quality of DNA was checked by 1% agarose gel electrophoresis and PicoGreen® dsDNA Assay (Invitrogen).
SureSelect sequencing libraries were prepared following the manufacturer’s instructions using a Bravo automated liquid handler. One ug of genomic DNA in 120 mL EB buffer was fragmented to a median size of 150 bp using a Covaris-S2 (Covaris) with the following settings: duty cycle 10%, intensity 5, cycles per burst 200, and mode frequency sweeping for 360 s at 4°C. Capillary electrophoresis on DNA 100 chips (Bionalyzer, Agilent) was used to evaluate the efficiency of the fragmentation. Following the manufacturer’s protocol (Agilent), sequencing adapters were ligated on the DNA fragments. PCR was used to amplify the adapter ligated DNA. Capillary electrophoresis (Bioanalyzer, Agilent) was used to ensure the quality of the PCR products. In preparing the hybridization buffer, #1, #2, #3, and #4 reagents (Agilent) were mixed. The amplified DNA fragments were concentrated to 500 ng in 3.4 ul. The 500 ng of DNA was mixed with SureSelect block #1, #2, and #3 reagents (Agilent). The hybridization buffer and DNA blocker mix were incubated for 5 min at 95°C followed by 10 min incubation at 65°C in a thermal cycler. Rnase block (Agilent) was added to the SureSelect oligo capture library (Agilent) and the capture library was incubated for 2 min at 65°C. In a thermal cycler, the hybridization buffer followed by the DNA blocker mix was added to the capture library and the mixture was incubated for 24 hours at 65°C. Fifty ul of streptavidin coating the Dynal MyOne Streptavidin T1 (Invitrogen) were washed three times with 200 ml SureSelect binding buffer (Agilent) and resuspended in 200 ml of the binding buffer. After being added to the bead suspension, the hybridization mixture was incubated for 30 min at room temperature with mixing. The beads were washed with 500 mL SureSelect wash buffer #1 (Agilent) for 15 min at room temperature followed by
three times wash with 500 mL SureSelect buffer #2 (Agilent) for 10 min at 65°C and DNA was eluted with 50 mL SureSelect elution buffer (Agilent) for 10 min at room temperature. Fifty mL of SureSelect neutralization buffer (Agilent) was added to the eluted DNA. Purification of the reaction product was done with the AMPure XP bead (Beckman). Using Herculase II Fusing DNA Polymerase (Finnzymes), the captured library was amplified to add index tags, capillary electrophoresis (Bioanalyzer, Agilent) then being used to verify the quality of the amplified libraries.

The 6 libraries, index tagged in equimolar amounts in the pool, were combined after GPCR using SYBR Green PCR Master Mix (Applied Biosystems). Cluster generation appeared in the flow cell on the cBot automated cluster generation system (Illumina) and the flow cell was loaded on the HISEQ 2000 sequencing system (Illumina) for sequencing with 2x100bp read length.

6. Whole Exome Sequencing and variant analysis

On average, 5.96 gigabases of raw sequence were generated per sample to achieve an average of 53.18× coverage of the WES target regions (51 megabases). The 6 individuals’ raw sequencing data was screened for common artifacts prior to comparison with control data sets. SNP variants were selected and filtered according to the following criteria. Variants were considered to be common if present in the 1000 Genomes Project East Asian database. Variants present in 1000 Genomes Project, dbSNP137, and 126 Korean Population data were filtered out from case variants. Variants were eliminated if they had
minor allele frequencies (MAF) $\geq 0.05$ from all control datasets used for comparison. Effect, impact (high or moderate), and loss of function in WES were established with SnpEff v3.3h (http://snpeff.sourceforge.net/).

B37 was used to build the reference genome. For prediction of variants (variant calling), only reads mapping to a unique position in the reference genome were used. Variants were identified with the Genome Analysis Toolkit (GATKv2.7-1) software, taking into account the single nucleotide polymorphisms obtained from the Single Nucleotide Polymorphism Database (dbSNP, National Center for Biotechnology Information) and from the 1000 Genomes project. Annotated, non-synonymous variants found in affected individuals were compared to variants present in the non-affected relatives. Variants present in affected individuals but not in healthy individuals were ranked based on this analysis to generate a list of candidate genes. The filtering process, which involved further bioinformatics analysis, is described in Figure 1.
The filtration and prioritization framework used for data analysis

**Fig. 1.** The filtration and prioritization framework used for data analysis
7. Statistical Analysis

Sequence Kernel Association Test (SKAT) was implemented to investigate candidate variants. SKAT combines squared single-variant score statistics, making it robust in its inclusion of neutral and protective variants and more powerful than pooling tests. SKAT is a flexible, computationally efficient regression approach that tests for association between variants in a region (both common and rare) and a dichotomous (i.e. case-control study) or continuous phenotype while adjusting for covariates, such as principle components, to account for population stratification. SKAT, which performs region-based testing, directly conducts multiple regressions of a phenotype on genotypes for all variants in the region, adjusting for covariates. SKAT results including magnitudes and directionality of the associations are based solely on estimates made from actual data sets. SKAT results were categorized by p-value.

8. Gene set enrichment analysis

GSEA (DAVID Bioinformatics Resource 6.7) was performed to determine the statistical significance of gene sets recovered from SKAT analysis. DAVID (Database for Annotation, Visualization, and Integrated discovery) is a web-accessible program that integrates functional genomic annotations with intuitive graphical summaries. With the GSEA results, cluster enrichment analysis was performed to build a protein functional network. We used following categories in DAVID. In “Gene Ontology” section:
“GOTERM BP ALL”, “GOTERM MB ALL”. In “Protein Domains” section: “INTERPRO”, “SMART”. We used the DAVID v6.7 service to compute functional enrichment for genetic variants from implant clustering failure patients (http://david.abcc.ncifcrf.gov/) \(^{25,26}\). The genetic variants from implant clustering failure patients contain 2,022 unique proteins. DAVID recognized 2,013, which were used in subsequent DAVID functional analyses. We used terms with p-value calculated after Benjamini-Hochberg correction less than 0.05.

9. Construction of protein functional network

Protein functional network in Figure 2 was built based on the top 10 cluster with highest enrichment scores. We linked two functional terms if they shared more than five proteins. Nodes represent enriched clusters of gene functions in cell morphology (green), cell adhesion (orange), and regulation of metal ion concentration (purple). Size of nodes shows the number of genetic variants in each functional group. Edge thickness is proportional to number of shared genes. Functional modules were manually grouped and labeled using Cytoscape 2.8 \(^{27}\) (www.cytoscpace.org).
III. Results

The mean age of the 6 individuals tested for WES was 59.17±7.78. A total of 55 implants were placed and 26 implants explanted from selected individuals (Table I). Quality Control (QC) run on saliva samples obtained from the six individuals confirmed samples were adequate for WES.

1. Whole Exome Sequencing and variant filtering findings

An average of 59,053,587 reads and 5,694 megabases were obtained from the six individual’s WES results. An average of 2,886,696,571 bases was aligned with mean coverage depth of 56.04. All information regarding number of reads, sample coverage and sequencing depth, as well as the data quality, are summarized in Table II. A total of 93,955 variants were established and SnpEff, which annotates variants to eliminate ones not altering protein sequence and predicts effects of each variant on genes, narrowed variants down to 24,543. The 1000 Genomes East Asian population and Korean randomized subsamples were also used to filter out common variants by eliminating ones with p-value greater than 0.05 from the two control groups. 9,243 variants were left after filtering based on the 1000 Genomes East Asian population and 5,056 after Korean randomized subsamples (Table III).

Variants were translated into genes, with a total of 3,724 gene transcripts remaining after filtering against 1000 Genomes East Asian population and Korean randomized
subsamples. SKAT analysis, which can increase statistical power when only a small number of cases are available, was applied to reveal statistical significance of all genes. Gene transcripts were categorized by p-value. 2,022 gene transcripts remained after ranking gene transcripts by p-value (p-value < 0.05).

**Table II.** Summary of number of reads and coverage

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Yield (Mbases)</th>
<th># Reads</th>
<th>Total bases aligned</th>
<th>Mean coverage depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp1_idx_2</td>
<td>5,428</td>
<td>53,743,622</td>
<td>2,637,087,120</td>
<td>51.20</td>
</tr>
<tr>
<td>exp2_idx_10</td>
<td>5,689</td>
<td>56,329,914</td>
<td>2,602,866,290</td>
<td>50.53</td>
</tr>
<tr>
<td>exp3_idx_3</td>
<td>7,206</td>
<td>71,350,078</td>
<td>3,458,057,942</td>
<td>67.14</td>
</tr>
<tr>
<td>exp4_idx_9</td>
<td>4,624</td>
<td>45,781,248</td>
<td>2,239,174,257</td>
<td>43.47</td>
</tr>
<tr>
<td>exp5_idx_11</td>
<td>5,964</td>
<td>59,049,854</td>
<td>3,138,909,381</td>
<td>60.94</td>
</tr>
<tr>
<td>exp6_idx_12</td>
<td>6,875</td>
<td>68,066,808</td>
<td>3,244,084,437</td>
<td>62.98</td>
</tr>
</tbody>
</table>

**Table III.** Number of genes common to the 6 affected individuals in each filtering step

<table>
<thead>
<tr>
<th>Filtering Step</th>
<th>Number of Variants</th>
<th>Number of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Gene Transcripts</td>
<td>93,955</td>
<td>19,919</td>
</tr>
<tr>
<td>Impact</td>
<td>24,543</td>
<td>9,651</td>
</tr>
<tr>
<td>1000 Genomes Project</td>
<td>9,243</td>
<td>4,677</td>
</tr>
<tr>
<td>Korean randomized subsample</td>
<td>5,056</td>
<td>3,724</td>
</tr>
</tbody>
</table>
2. Bioinformatics findings: GSEA and network analysis

Gene set enrichment analysis (GSEA) was applied to investigate genetic variants in groups of genes sharing common biological function, domain or pathway. We used the DAVID (DAVID v6.7) to discover enriched functional-related gene groups and corrected the results using the Benjamini-Hochberg corrected p-value to correct multiple comparison errors in gene sets that can arise in WES analysis. The DAVID provides a novel way to functionally analyze a large number of variants in a high-throughput fashion by classifying them into gene groups based on their annotation term co-occurrence. 927 gene sets were annotated and 175 gene sets remained after Benjamini-Hochberg corrected p-value less than 0.05. Multiple domains emerged and a few gene sets were looked into based on their roles in biological process. GSEA detected multiple gene sets within the cell adhesion, metal ion binding, and cytoskeleton.

The protein functional network was used for further analysis to see how genes from different domains react with each other in a biological system and how interactions between genes might affect pathogenesis of peri-implantitis. To examine the function of genetic variants from peri-implantitis patients, we searched for significantly enriched gene function clusters. Cluster enrichment was performed for the whole list of genes with mutations and the top 10 enriched clusters were picked for network visualization (Table IV). Each cluster, terms and genes in protein functional network analysis were described using circular shape, nodes and line (Figure 2). As nodes describe terms with p-value after Benjamini-Hochberg corrected p-value of less than 0.05 and line expressed number
of genes shared by each terms, size of nodes shows number of genetic variants and thickness of link describes number of shared genes. Nodes represent enriched clusters of gene functions in cell morphology, cell adhesion, and regulation of metal ion concentration.

Network analysis showed metal ion binding in the middle of all clusters and terms implying that ion binding is shared by genes between cell adhesion and cytoskeleton. Cell adhesion and cytoskeleton clusters, which affect cell morphologies showed mutations. In addition, cytoskeleton cluster has functional role for cell morphology, such as dynein, and Myosin, including Actin and ATPase activity. Clusters shared multiples genes through the protein functional network and all clusters showed interactions through metal ion binding.

3. Known peri-implantitis genes

Genes which are known to be associated with peri-implantitis and periodontitis were tested for their presence in our case group; IL-1A, IL-1B, and TNF all appeared in our data set (Table V).
Table IV. Top 10 clusters ranked by enrichment score from GSEA results

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Enrichment Score</th>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>Benjamini</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadherin</td>
<td>13.69</td>
<td>IPR013164: Cadherin, N-terminal</td>
<td>32</td>
<td>1.66</td>
<td>4.40E-13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR002126: Cadherin</td>
<td>44</td>
<td>2.29</td>
<td>4.40E-13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR015492: Protocadherin gamma</td>
<td>18</td>
<td>0.94</td>
<td>1.17E-12</td>
</tr>
<tr>
<td>Cell adhesion</td>
<td>12.64</td>
<td>GO:0007155: Cell adhesion</td>
<td>139</td>
<td>7.23</td>
<td>2.73E-11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GO:0022610: Biological adhesion</td>
<td>139</td>
<td>7.23</td>
<td>1.55E-11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GO:0016337: Cell-cell adhesion</td>
<td>65</td>
<td>3.38</td>
<td>1.77E-07</td>
</tr>
<tr>
<td>EGF</td>
<td>5.42</td>
<td>IPR013032: EGF-like region, conserved site</td>
<td>67</td>
<td>3.49</td>
<td>1.87E-07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR006210: EGF-like</td>
<td>47</td>
<td>2.45</td>
<td>4.10E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR000742: EGF-like, type 3</td>
<td>43</td>
<td>2.24</td>
<td>4.95E-04</td>
</tr>
<tr>
<td>Actin</td>
<td>3.59</td>
<td>IPR001715: Calponin-like actin-binding</td>
<td>21</td>
<td>1.09</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR018159: Spectrin/alpha-actinin</td>
<td>12</td>
<td>0.62</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SM00033: CH</td>
<td>21</td>
<td>1.09</td>
<td>0.02</td>
</tr>
<tr>
<td>Metal ion binding</td>
<td>3.58</td>
<td>GO:0043167: Ion binding</td>
<td>513</td>
<td>26.69</td>
<td>5.65E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GO:0046872: Metal ion binding</td>
<td>500</td>
<td>26.01</td>
<td>7.09E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GO:0043169: Cation binding</td>
<td>504</td>
<td>26.22</td>
<td>7.26E-04</td>
</tr>
<tr>
<td>Laminin/lectin</td>
<td>3.40</td>
<td>IPR001791: Laminin G</td>
<td>15</td>
<td>0.78</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPR013320: Concanavalin A-like lectin/glucanase</td>
<td>19</td>
<td>0.99</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SM00282: LamG</td>
<td>15</td>
<td>0.78</td>
<td>0.02</td>
</tr>
<tr>
<td>Term</td>
<td>Count</td>
<td>%</td>
<td>Benjamini</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------</td>
<td>------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0032559~adenyl ribonucleotide binding</td>
<td>210</td>
<td>10.93</td>
<td>2.81E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0030554~adenyl nucleotide binding</td>
<td>217</td>
<td>11.29</td>
<td>6.06E-04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GO:0005524~ATP binding</td>
<td>205</td>
<td>10.67</td>
<td>5.67E-04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>Benjamini</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR003961:Fibronectin, type III</td>
<td>37</td>
<td>1.93</td>
<td>0.02</td>
</tr>
<tr>
<td>IPR008957:Fibronectin, type III-like fold</td>
<td>33</td>
<td>1.72</td>
<td>0.09</td>
</tr>
<tr>
<td>SM00060:FN3</td>
<td>37</td>
<td>1.93</td>
<td>0.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>Benjamini</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR013602:Dynein heavy chain, N-terminal region 2</td>
<td>8</td>
<td>0.42</td>
<td>0.04</td>
</tr>
<tr>
<td>IPR004273:Dynein heavy chain</td>
<td>8</td>
<td>0.42</td>
<td>0.04</td>
</tr>
<tr>
<td>IPR011704:ATPase associated with various cellular activities</td>
<td>8</td>
<td>0.42</td>
<td>0.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>%</th>
<th>Benjamini</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR000048:IQ calmodulin-binding region</td>
<td>23</td>
<td>1.20</td>
<td>0.01</td>
</tr>
<tr>
<td>IPR001609:Myosin head, motor region</td>
<td>14</td>
<td>0.73</td>
<td>0.01</td>
</tr>
<tr>
<td>SM00242:MYSc</td>
<td>14</td>
<td>0.73</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Fig. 2. Analysis of gene function enrichment and construction of functional network
Gene functions significantly enriched in genetic variants from implant clustering failure patients are visualized as a functional cluster network. Nodes represent enriched clusters of gene functions in cell morphology (green), cell adhesion (orange), and regulation of metal ion concentration (purple). Size of nodes shows the number of genetic variants in each functional group. Edge weight is proportional to the number of shared genes. Clusters of functionally related nodes were manually circled and labeled.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Transcript</th>
<th>Variant Sample Position</th>
<th>Samples Affected</th>
<th>Samples Heterozygous</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL6</td>
<td>ENST00000404625</td>
<td>exp1_idx_2:7:22771039</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL1B</td>
<td>ENST00000416750</td>
<td>exp6_idx_12:2:113590977</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL1B</td>
<td>ENST00000418817</td>
<td>exp6_idx_12:2:113590977</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL1B</td>
<td>ENST00000432018</td>
<td>exp6_idx_12:2:113590977</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL1B</td>
<td>ENST00000263341</td>
<td>exp6_idx_12:2:113590977</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL1A</td>
<td>ENST00000263339</td>
<td>exp3_idx_3:2:113537072</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL6</td>
<td>ENST00000258743</td>
<td>exp1_idx_2:7:22771039</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL6</td>
<td>ENST00000407492</td>
<td>exp1_idx_2:7:22771039</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TNF</td>
<td>ENST00000449264</td>
<td>exp5_idx_11:6:31543574</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IL6</td>
<td>ENST00000401630</td>
<td>exp1_idx_2:7:22771039</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
IV. Discussion

With the development of different types of genome sequencing and associated analytical tools, biomarker discovery has emerged as a major area of research throughout medicine. In 2009, four individuals with Freeman-Sheldon syndrome (FSS), a rare autosomal dominant disorder, were exome-sequenced to prove that such sequencing could identify causal genetic variants. As WES is known to sequence thousands of functional genes at a time, it has become the tool of choice for discovering causative genetic variants, especially in Mendelian diseases. Although exome sequencing is often used to investigate Mendelian disorders, it also expands our knowledge of novel mutations of established genes linked to a particular disorder such as celiac disease and helps to uncover the complex interplay between modifier variants that contributes to a disease phenotype. For these reasons, WES has been widely used to discover genetic factors related to a range of diseases in various medical fields, particularly for diseases that exhibit broad genetic or phenotypic heterogeneity.

The WES approach held promise for our current research given the rarity and distinct phenotype of implant clustering failure in peri-implantitis. There were following limitations in current study: small number of case population to generalize the trend in whole Korean population, to link pathogenesis of peri-implantitis solely to genetic causalities because peri-implantitis is a multi-factorial diseases, and use of saliva sample over blood sample. Saliva sample collection method was chosen based on previous studies confirming its quality suitable for genetic sequencing and high percentage of case
recruitment\textsuperscript{33,34}. With careful consideration, WES approach was taken for current study. WES results from this study showed a whole set of variants, necessitating filtering to discover genetic variants specifically related to peri-implantitis. WES annotation from the 6 individuals in this study revealed a number of variants for further analysis. Standard methods used to test a disease such as peri-implantitis for association with a single common genetic variant are underpowered given the insufficient sample or effect sizes\textsuperscript{35,36}. Hence, the sequence kernel association test (SKAT) was used to accommodate our sample size and to improve statistical power.

SKAT result was further taken for GSEA study via DAVID and various domains were discovered.

Adhesion between cells and implant in the long term involves regulation of aspects of cell expression such as cell membrane proteins, ECM proteins, integrins and cytoskeleton proteins, all which work together to maintain the proper adhesion. The external faces of focal contacts present specific receptor proteins. Cadherins, which showed highest association in this study, are a class of type-1 transmembrane proteins which play an important role in cell adhesion, forming adherence junctions to bind cells within tissues together and also participate in implant-bone adhesion. ECM in intracellular fluid is another important component for cell adhesion. Integrins are a major family of cell surface-adhesion receptors that can mediate cell-cell, cell-matrix and cell-pathogen interactions. Most integrins are not intrinsically active and often expressed on cell surface in an inactive state, in which they neither bind ligands nor signal. Metal ion Ca\textsuperscript{2+} play an important role in keeping integrins in inactive state and removal of Ca\textsuperscript{2+} or addition of Mn\textsuperscript{2+} will strikingly increase binding affinity and
adhesiveness of almost all integrins. Integrin-mediated adhesion and signaling events are important in normal physiological responses, immune response, tissue morphogenesis, wound healing, hemostasis, cell survival and cell differentiation. Conversely, dysregulation of integrins are involved in the pathogenesis of many diseases, including cancer metastasis, auto-immune disease and thrombotic vascular diseases. It may be assumed that insufficient metal ion concentration will cause to decrease cell to titanium adhesion during the healing stage of implant resulting inflammatory disease. Another fact is that some integrins are able to bind several ligands such as laminin and fibronectin. The FN domain, a major glycoprotein in the extracellular matrix, binds specifically to titanium implants, which can serve as a ligand for a dozen members of the integrin receptor family. It was reported that FN III 7-10 and FN III 9-10 synthetic peptides are effective in osteoblast adhesion, necessary for successful dental implant outcome. It was also reported that FN coatings on titanium implants were advantageous for peri-implant bone formation. Another high-ranking gene cluster was the Epidermal Growth Factor (EGF) domain, is critical in stimulating cellular proliferation, differentiation, and survival. Consolaro et al concluded that EGF in the saliva and in the epithelial cells stimulates peri-implant epithelial proliferation, thereby triggering the formation of the peri-implant junctional epithelium. It is remarkable that all high ranking domains in the study play a critical role in adhesion of cells to titanium surface and in the development of peri-implantitis. In the last ten years, implantology research has focused on the analysis of peri-implant fluids with the basic aims of identifying potentially valid biochemical and immunological markers of inflammatory processes and their levels and/or of predicting risk for the
onset of peri-implant disease. There have been several genetic studies of the role of cytokines in peri-implantitis. Cytokines are hormonal regulators or signaling molecules of host responses to infection, immune responses, inflammation, and trauma, making them significant in peri-implantitis, which results from an unregulated host inflammatory response to antigen bacterial determinants from dental plaque. Clinical research indicates that monitoring the dynamics of local cytokine levels during peri-implantitis, together with research into gene polymorphism for these cytokines and other genes involved in the inflammatory process, is a valid means of achieving new methods of diagnosis, prognosis and treatment of peri-implantitis. A recent systematic review on the association between genetic predisposition and dental implant biological complications implied that there is no strong association between specific genetic polymorphism (IL-1, IL-2, IL-6, TNF-α or TGF-β1) and dental implant failure in terms of biological complications although a there was a potential link between the IL-1 genotype and peri-implantitis. Genes known to be associated with peri-implantitis were also tested for our case group but did not show significant results, suggesting the existence of undiscovered variants linked with peri-implantitis.

The current study chose six patients with implant clustering failure to investigate genetic variants involved with peri-implantitis. Implant clustering failure is a technical term used to describe the phenomenon in which a few individuals have a concentrated risk for multiple implant failure and subsequently experience multiple losses. Among individuals sharing relatively similar environmental factors, implant clustering failure has low occurrence among the total population, suggesting its strong association to genetic factors rather than behavioral ones such as smoking, stress, or maintenance of dental
implant after surgery. The patients selected for the current study had severe peri-
implantitis leading to implant failure with the number of occurrences varying among the
individuals. Current research had limitations in selecting case population with the
exclusion of environmental factors known to affect peri-implantitis. Under similar
environmental influence, the distinctive phenotypes of the six patients with implant
clustering failure were very rare among implant patients, indicating that such clustering
failure was likely to be caused by individual genetic differences. With limitations in mind,
case population with implant clustering failure was chosen for genetic approach and WES
was thus applied in the current study to detect pathogenic genes associated with peri-
implantitis in these patients experiencing implant clustering failure.

Other risk factors involved with peri-implantitis include smoking, stress, diabetes,
osteoporosis, and genetics. Because multiple causalities are involved in peri-implantitis,
patient recruitment for current study was very difficult. Environmental factors such as
smoking behavior can greatly affect the occurrence of peri-implantitis and careful
consideration to eliminate group of patients with environmental risk factors was taken
for current study. Also, randomized subsample data of 126 Koreans was chosen with
exclusion of disease that can affect peri-implantitis such as diabetes and osteoporosis
but monitoring of other risk factors in this randomized subsample group was limited.
Machalowicz et al indicated that genetics constituted the most important factor
influencing differences in periodontal disease, based on studies involving twins. Also,
this group mentioned individuals vary from one another because of differences in
genetic makeup and environments implying the variances in the population for a given
measure can be partitioned into genetic and environmental variances. Current study focused on choosing individuals experiencing severe peri-implantitis with relatively similar environmental factors and showing distinctive phenotype to perform genetic association study.

Limited number of study population recruitment and difference in arrangement of age between 126 Koreans from the GSK project and study population restricted concrete conclusion of the study. To fully elucidate the effect of environmental and genetic variances in peri-implantitis, WES results of each case population could be compared to family members without the indication of peri-implantitis.
V. Conclusion

Previous analyses of peri-implant fluids aimed at identifying potentially valid biochemical and immunological markers of inflammatory processes have focused on cytokines. However, our results suggest that various genes and gene sets related to factors involved in cell adhesion such as cadherin, fibronectin, integrins, EGF domains, and cytoskeletons play critical roles in the osseointegration and pathogenesis of peri-implantitis. Interestingly, these two gene sets are indirectly linked via the metal ion binding protein. One may conclude that regulatory imbalance in metal ion concentration elicits dysfunction in cell morphology and cell adhesion, eventually causing peri-implantitis. Discovery of metal ion in protein functional network study implied dysregulation of integrins which can have impact on surface adhesion would affect the occurrence of peri-implantitis. Genetic diagnosis before implant surgery would highlight the genes related to peri-implantitis and help clinicians to provide appropriate treatment to those patients in the future. Also, pathogenesis of peri-implantitis can be investigated through genetic research similar to current research and can generate treatment options for patients. Further research with more cases and controls, along with functional animal studies may yield legitimate biomarkers for early diagnosis of peri-implantitis. Data from this research will be used to develop analytic methods for small case-only samples.
References


국문요약

임플란트 군집성 실패 환자로부터 임플란트 주위염에서의 발병 유전자의 조사:

Whole-exome sequencing 연구

(지도: 한 동 후 교수)

연세대학교 대학원 치의학과

이 수 형

목적: 임플란트 주위염은 다양한 환경적, 유전적 인과관계와 함께 복합적인 원인에 의해 혼하게 발생하는 치주염증이며, 아직 확실한 병인은 없는 것으로 알려져 있다. 비교적 유사한 환경적 요소를 가진 환자들 사이에서 임플란트 주위염의 정도의 차이는 질병의 병인을 이해하고 조절하기 위하여 조사되어져야 하는 유전적인 면을 제시한다. 본 연구의 목적은 전체 액송 시퀀싱, 유전자 군 인리치먼트 분석, 그리고 네트워크 분석을 통해서 임플란트 주위염의 발병과 관련된 원인 유전자를 찾아내는 것이다.
방법: 심한 임플란트 주위염으로 인하여 다수의 군집성 임플란트 실패를 가진 서로 연관성이 없는 6 명의 환자들이 본 연구를 위하여 선택되었다. 환자군은 임플란트 주위염에 영향을 주는 환경적인 인과관계가 최소인 비교적 건강한 생활방식을 가졌으며, 동일한 환자 내에서 다수의 임플란트 소실을 보임으로써 유전적 원인이 크게 작용한 것으로 생각되었다. 임플란트 주위염을 야기하는 유전자들을 조사하기 위하여 유전적 및 생물 정보학적 접근법을 시행하였다. OG-500(DNAgenoTeK)을 통해 모아진 타액 샘플로 전체 엽서 시퀀싱을 시행하였고, 모든 일반 대조군 자료로부터 0.05 이상의 이상 유전자 간 대립 반도를 갖는 공통 유전적 변이들을 제거하였고 중등도 이상의 영향력과 기능의 상실을 갖는 변이들을 비교에 사용하였다. 유전적 변이들과 연관된 기능 군(functional group)들을 밝혀내기 위하여 유전자 군 인리치먼트 분석을 시행하였다. 기능 군 간의 관계를 찾아내기 위하여 네트워크 분석을 시행하였다.

결과: dbSNP, 1000 Genomes East Asian population 그리고 건강한 한국인의 우작위적 부표분 자료(GSK project)를 이용한 필터링 후 2,022 개의 유전자가 남았다. 927 개의 유전자 군들 중 0.05 미만의 p-value 를 갖는 175 개의 유전자 군들이 유전자 군 인리치먼트 분석(DAVID)을 통해 얻어졌다. 클러스터 인리치먼트를 통해 상위 10 개의 클러스터를 선정하였으며, 이들은 세포 골격, 세포 접착, 그리고 금속 이온 결합에 중요도를 보여주었다. 기능 군들
내에서, 금속 이온 결합은 모든 클러스터들의 가운데에 위치하였으며, 이는 금속 이온 농도 조절의 기능 이상은 임플란트 실패로 이어지는 세포 형태 또는 세포 접착에 영향을 줄 수 있다는 것을 나타낸다.

결론: 본 연구의 결과는 cadherin, fibronectin, EGF domain 과 같은 세포 접착 그리고 세포 골격과 연관된 다양한 유전자들과 유전자 그룹이 골유착과 임플란트 주위염의 발병에 중요한 역할을 하고 있음을 제시한다. 흉미롭게도, 이 두 유전자 군은 금속 이온 결정 단백질을 통해 간접적으로 연결되어져 있다. 이는 금속 이온 농도 조절의 브라켓으로 인한 세포 형태와 세포 접착에서의 기능 이상과, 금속 이온 결정에 의하여 조절되며 세포의 표면 접착에 중요한 역할을 하는 integrin의 조절장애가 임플란트 주위염의 발병에 영향을 줄 수 있다는 결론을 내릴 수 있다.

핵심되는 말: 임플란트 주위염, 발병 유전자, 임플란트 군집성 실패, 전체
액소 시편성, 유전자군 인리치먼트 분석, 네트워크 분석, 세포 접착, 세포 골격, 금속 이온 농도

39