Bioinformatic Prediction of Leader Genes in Human Periodontitis

Ugo Covani,*† Simone Marconcini,*† Luca Giacomelli,† Victor Sivozhelevov,†‡ Antonio Barone,* and Claudio Nicolini†‡

Background: Genes involved in different biologic processes form complex interaction networks. However, only a few have a high number of interactions with the other genes in the network. In previous bioinformatics and experimental studies concerning the T lymphocyte cell cycle, these genes were identified and termed "leader genes." In this work, genes involved in human periodontitis were tentatively identified and ranked according to their number of interactions to obtain a preliminary, broader view of molecular mechanisms of periodontitis and plan targeted experimentation.

Methods: Genes were identified with interrelated queries of several databases. The interactions among these genes were mapped and given a significance score. The weighted number of links (weighted sum of scores for every interaction in which the given gene is involved) was calculated for each gene. Genes were clustered according to this parameter. The genes in the highest cluster were termed leader genes.

Results: Sixty-one genes involved or potentially involved in periodontitis were identified. Only five were identified as leader genes, whereas 12 others were ranked in an immediately lower cluster. For 10 of 17 genes there is evidence of involvement in periodontitis; seven new genes that are potentially involved in this disease were identified. The involvement in periodontitis has been completely established for only two leader genes.

Conclusions: We applied a validated bioinformatics algorithm to increase our knowledge of molecular mechanisms of periodontitis. Even with the limitations of this *ab initio* analysis, this theoretical study can suggest ad hoc experimentation targeted on significant genes and, therefore, simpler than mass-scale molecular genomics. Moreover, the identification of leader genes might suggest new potential risk factors and therapeutic targets. *J Periodontol 2008;79:1974-1983.*

KEY WORDS

Computational biology; genomics; periodontitis.

* Unit of Oral Pathology, School of Dental Medicine, University of Genoa, Genoa, Italy.
† Interuniversities Research and Didactic Service Center on BioOrganic Nanosciences and Nanotechnologies, University of Genoa and Nanoworld Institute, Genoa, Italy.
‡ Electronics Biotechnology Advanced Foundation, Rome, Italy.

This is a 2010 revision of the Covani et al. paper originally published in the October 2008 issue. The revision is in response to issues that are detailed in an erratum (*J Periodontol* 2010;81:1237).

When the human genome project was completed and published, a new era of human genetics was born.[1] This resource now permits the analysis of genes and their variants in any number of human conditions. The potential to associate gene variations with disease and disease susceptibility is highly attractive.

Periodontitis is a chronic infectious disease of the supporting tissues of the teeth. Because of bacterial infection, the periodontal tissues become inflamed and are slowly destroyed by the action of the inflammatory process. If left untreated, the teeth lose their ligamentous support to the alveolar bone, become mobile, and are eventually lost.

A large majority of periodontitis patients respond well to conventional therapies.[2] However, the results of some studies have suggested that a small percentage of patients present a poor response.[3,4] Those patients are defined as "downhill patients" (loss of four to nine teeth) or as "extremely downhill patients" (loss of 12 to 23 teeth). In a study of 600 patients conducted by Hirschfeld and Wasserman in a private periodontal practice, the percentages of downhill patients and extremely downhill patients were 13% and 4%, respectively, after a 22-year administration of a periodontal treatment.[3] Similar findings were observed in another study of 100 patients

doi: 10.1902/jop.2008.080062

conducted by McFall, in which the percentages of downhill patients and extremely downhill patients were 15% and 8%, respectively, after a 15-year follow-up.[4] Of note, even if the patients in the extremely downhill subgroup were only a small percentage of the total population, the largest number of tooth losses occurred in this group.[3,4] On these bases, downhill and extremely downhill patients are often considered as refractory to the conventional treatment. Moreover, it should be considered that the classification of periodontal diseases and conditions defined as "refractory" both refractory chronic periodontitis and refractory aggressive periodontitis patients.[5]

Genetic risk factors have been proposed to influence the natural history of periodontitis.[6-9] The presence of a genetic risk factor directly increases the probability of periodontal disease development.

Periodontitis is a complex multifactorial disease. These kinds of diseases (e.g., Alzheimer's disease, Crohn's disease, and cardiovascular diseases) usually present a relatively mild phenotype and are slowly progressive.[10] The pathophysiology of complex diseases is characterized by various biologic pathways. Complex diseases are associated with variations in multiple genes, each providing a small overall contribution to the pathologic process: complex diseases are typically polygenic.[11] Therefore, the knowledge of molecular mechanisms of these diseases must deal with a large number of genes, instead of a few or even only one gene.[10] Moreover, genes are interacting with one another, forming complex networks. Interactions between genes may be direct (physical interactions between the proteins, confirmed by experimental techniques, such as nuclear magnetic resonance or crystallography) or indirect (involvement in the same metabolic pathway or coexpression in different conditions).[12]

In the past 2 decades, many genes that were implicated in simple (Mendelian) diseases have been identified by using genetic linkage and positional cloning methods.[13] Although these methods have been remarkably successful in identifying high relative risk genes, they have not been successful in identifying genes involved in complex diseases. This failure is the result of three main features. First, complex diseases are more likely to be caused by several, and even numerous, genes, each with a small overall contribution and relative risk. Therefore, the disease genes in complex diseases are considered disease-modifying genes.[11] Second, complex diseases typically vary with regard to the severity of symptoms and age of onset, which results in difficulty in defining an appropriate phenotype and selecting the best population to study. Third, etiologic mechanisms may vary, involving several biologic pathways.

Genomic analysis of complex disease may rely upon some appropriate experimental methodologies. Microarray technology, developed in the 1990s, allows the profiling of a whole genome via the analysis of thousands of transcripts in a single experiment.[14] In this way, researchers may obtain a deeper view and understanding of the interactions and expression levels of several thousands of genes simultaneously.[14] The use of microarray technology resulted in important improvements both to basic and applied research and, potentially, may become a turning point in the practice of medicine.[14] Microarrays measure the relative gene expression of different conditions, such as normal versus pathologic tissues. For instance, a recent study[15] used microarray technology to identify gene expression changes in downhill subjects. Moreover, microarray may potentially play a role in the molecular diagnosis of different diseases and in the prediction of drug efficacy and safety in different individuals.[14] New, and maybe even more promising, results might be obtained from the use of protein arrays.[16] However, the huge amount of data coming from microarray experiments may often raise experimental complications and difficulties in the analysis. Moreover, most genes displayed on an array are often not directly involved in the cellular process being studied. Commercial arrays with a lower number of genes, usually 150 to 200, are available, but the genes displayed are usually chosen without a precise consideration of the particular target of the study.[17] Therefore, it may be difficult to fully understand interactions among genes involved in a given process by relying only upon microarray analysis. Bioinformatics may aid in the understanding of microarray-derived data. Sophisticated software has been written to map gene interactions from microarray data,[18] which has resulted in the accumulation of a large amount of data on gene interaction (~800,000 interactions between genes are known). However, examples of focused searches in databases with statistics and data-mining methods are scant.[18]

Recently, we proposed the leader gene approach, a search/statistics algorithm that gave promising results when applied to the human T lymphocyte cell cycle and osteogenesis.[18,19] This bioinformatics *ab initio* method is based on the systematic search for genes involved in a given process and on their ranking according to the number of all experimentally established interactions, as derived from free Web-available databases, such as STRING (Search Tool for the Retrieval of Interacting Genes, Heidelberg, Germany).[12] Genes belonging to the highest rank are defined as "leader genes" because they may be assumed to play an important role in the analyzed processes. Therefore, the leader gene approach can suggest a list of the few most relevant genes in a given cellular process, according to the already available experimental data, and may be useful in interpreting the microarray

experiments and in guiding targeted clinical experimentations.[20] The application of this algorithm to the human T lymphocyte cell cycle invariably identified only six genes, among the several hundreds involved, with a significantly higher number of interactions compared to the others.[18] The same six leader genes are involved in the cell cycle control at important progression points. The identified leader genes are actually involved in different key points of the cell cycle. These results were fully confirmed in a targeted experimental analysis[19] using microarrays and a new and simple technology to acquire microarray images, the DNA analyzer.[21] However, a more detailed analysis in kidney transplants, in rats and in human patients, revealed that only the proper combination of experimental (microarray-based) and theoretical results (*ab initio* research of key genes based on data mining) gave an informative picture of the molecular mechanisms underlying such complex phenomena.[22]

In this study, we preliminarily applied the leader gene approach, only to perform an *ab initio* data-mining identification of potential leader genes in human periodontitis, to obtain a broader understanding of the molecular basis of this disease and to plan new ad hoc experimentations, which will be expected to provide significant findings. We also relate our results to present knowledge in periodontitis.

MATERIALS AND METHODS

The *ab initio* leader gene approach has been described in detail.[18] The bioinformatics/statistics algorithm followed is summarized in Figure 1.

First, the key genes involved in periodontitis were identified by an iterative search of large-scale gene databases. In particular, several search strategies were implemented and iteratively repeated until convergence. We performed our research in PubMed, GeneBank, Geneatlas,[23] and Genecards,[24] using pertinent key words chosen by experts, such as periodontitis or gingival inflammation, and MeSH terms, as well as all of their possible Boolean logic–based combinations. To avoid possible bias due to different nomenclature systems, we used official HUGO nomenclature. Only human genes were considered. In this way, it was possible to identify a preliminary set of genes representing every gene with an established role in periodontitis.

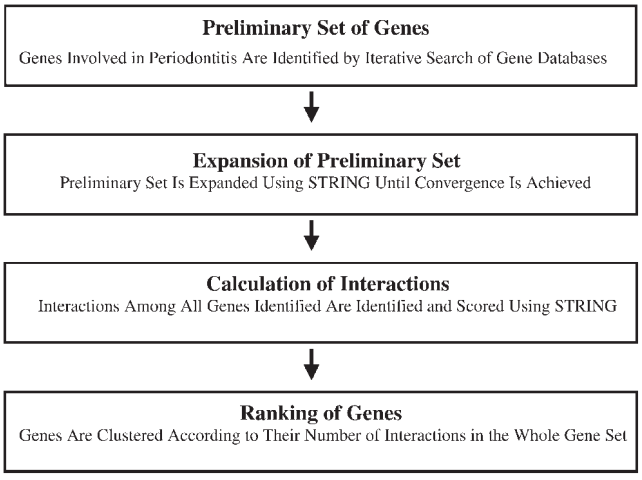

Figure 1.
Flow chart of the leader gene approach.

Then the preliminary set of genes was expanded using the Web-available software STRING (version 6.3), considering only direct interactions (i.e., physical contact between encoded proteins, gene expression microarray data, or direct linkage in the same pathway), with a high degree of confidence (>0.9; confidence values in STRING range between 0 and 0.99, with 0.99 being the highest confidence).[12] In this way, it is possible to identify new genes directly linked to those with an already established role in periodontitis and, therefore, potentially involved in this disease. Results were filtered using a further search in PubMed to discard false positives. The process was repeated until no new gene potentially involved in periodontitis was identified.

Then, an interaction map among identified genes was calculated using STRING. This software can give a combined association score to each interaction, representing the degree of confidence for each interaction. For every gene identified, we summed combined association scores, thereby obtaining a single score, named "weighted number of links."

Then genes were clustered, using hierarchical or K-means algorithms,[25,26] according to their weighted number of links. The genes belonging to the highest rank are termed leader genes; these genes have the highest weighted number of links compared to the others. Therefore, they may be assumed to have an important role in periodontitis. The other ranks are termed class B genes, class C genes, class D genes and so on, according to their importance. Genes with no identified interactions (i.e., weighted number of links = 0) are defined as orphan genes.

Differences among various classes were evaluated statistically using an analysis of variance test, with a Tukey-Kramer post hoc test. Statistical significance was set at $P<0.001$ to ensure a high level of data reliability.

To compare our findings to present knowledge in periodontitis, we searched for the number of citations in PubMed for each gene in relation to this disease, using a key word–based query. Research was last updated on July 3, 2007.

Lastly, for genes identified as leader or class B, we defined and calculated a specificity score parameter, i.e., number of interactions in periodontitis gene set/total number of interactions in the whole STRING database.

RESULTS

The first key words–based query in databases resulted in 37 genes with an established involvement in periodontitis. This preliminary set was expanded five times via STRING, until it reached convergence. Once convergence was reached, the expanded data set included 61 genes involved or potentially involved in periodontitis. Figure 2 shows the final interaction map among this set of genes.

The weighted number of links for each gene in this dataset is represented in Figure 3. Of 61 genes, 54 have known interactions, whereas the remaining seven (MMP1, GNRH1, GSTM1, CARD15, FOXP3, CYP2E1, and VDR) are orphan genes, i.e., genes with no known interactions in the whole gene set.

Cluster analysis of the weighted number of links identified the same five genes belonging to the highest cluster, i.e., to be leader genes: NFKB1, CBL, GRB2, PIK3R1, and RELA (Fig. 3 and Table 1). We also identified 12 class B genes (Table 1).

The analysis of variance revealed a significant difference in the weighted number of links. In particular, the post hoc test revealed that leader genes had a significantly higher weighted number of links compared to

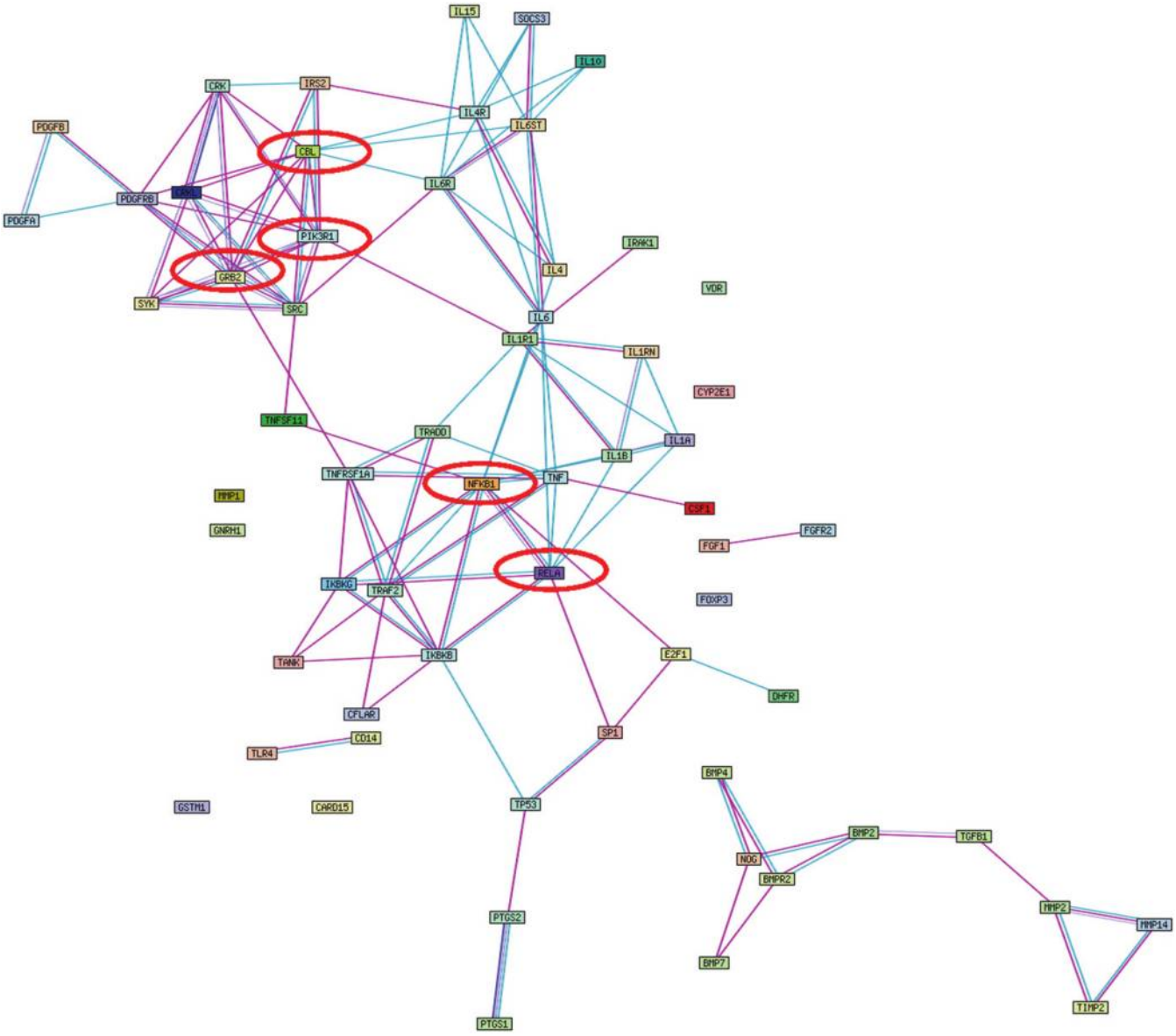

Figure 2.

Final map of interactions among genes involved in periodontitis according to STRING.[1,2] The lines that connect single genes represent physical interaction between proteins, confirmed by various experimental methods (magenta lines), correlation in gene expression experiments (dark blue), or involvement in the same metabolic pathway (light blue). Leader genes are circled in red.

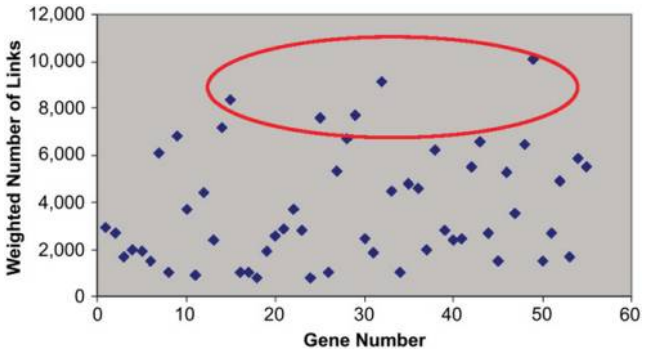

Figure 3.
Weighted number of links for each of the 61 genes involved in periodontitis. Genes are numbered according to a serial number assigned by the authors. Weighted number of links is the sum of combined association scores for every interaction among genes predicted with STRING software.[12] Genes belonging to the highest cluster, i.e., leader genes, are circled in red.

class B genes ($P<0.001$), and class B genes differed significantly from other classes ($P<0.001$ versus class C).

The search for citations of leader and class B genes concerning periodontitis revealed that only 10 of 17 genes had an already established involvement in this disease (Table 1). Among leader genes, RELA had two citations, whereas PIK3R1 had one. CBL, NFKB1d, and GBR2 had no citations. Conversely, class B genes were cited slightly more often. However, only TNF had a very high number of citations (242). Several genes not belonging to the leader gene class or class B are often studied in periodontitis. Table 2 shows the 10 genes with the highest number of citations with regard to this disease: only TNF belongs to class B; the remaining genes are included in lower important classes, and one of them (MMP1) is an orphan gene.

The specificity scores for each leader and class B gene are summarized in Table 1. Leader genes had a lower specificity score (mean: 9.4%) than class B genes (mean specificity: 17.2%).

DISCUSSION

Some articles[27,28] reviewed the strong suggestion that periodontitis has a genetic basis. Genetic and genomics research is rapidly increasing our understanding of the genetic basis of diseases and is being applied to develop new diagnostic and treatment strategies. Many diseases with dental, oral, and craniofacial manifestations have a genetic basis. Oral health status has recently been linked to a number of systemic conditions, including diabetes and cardiovascular disease.[29] Studies of these disease conditions suggest that multiple gene and gene–environment interactions are important determinants of susceptibility.[29] In this study, we used a data-mining approach to identify the genes involved or potentially involved in human periodontitis (61 genes) and to draw a map of interactions among them. We also ranked these genes on the basis of their number of interactions. We considered only direct interactions, i.e., those based on experimental observations described in the public domain and available in specific databases, such as STRING. Direct interactions include physical interactions between encoded proteins (e.g., ligand–receptor contact), gene expression data derived from microarray experiments, and proved involvement in the same metabolic pathways. We considered only interactions with a high degree of confidence in the STRING database, i.e., those with stronger experimental evidence. In this way, it was possible to limit database bias.

In particular, we identified a very small set of genes that showed the highest number of interactions, defined as leader genes, as suggested in previous studies.[18,22] The periodontitis leader gene group consisted of only five genes (NFKB1, CBL, GRB2, PIK3R1, and RELA). These genes may be assumed to play a major role in the control of periodontitis because their number of interactions was the highest in the whole gene set. It was also possible to identify 12 genes (class B genes) with an immediately lower importance in terms of interactions. Moreover, among 61 genes, seven were identified as orphans, i.e., they had no known interaction in this set of genes involved in periodontitis. To further characterize their precise role in this disease, the identification of orphan genes may be of particular importance in suggesting new experimentations and analysis.

Our analysis was conducted completely *ab initio*: we did not look at scientific literature when performing leader gene identification. The use of blindness may be an important proof of the validity of the method: after *ab initio* identification of leader genes, we searched the scientific literature to see if there was strong evidence (epidemiologic, clinical, or biochemical) for the involvement of leader and class B genes in periodontitis. However, if no evidence is found for a given gene, it might be important to verify if there are known, direct links to some other leader or class B gene playing an established role in periodontitis. In this case, a possible involvement in the disease may be assumed.

Experimental Evidence of Leader and Class B Gene Involvement in Periodontitis

Our bibliographic research revealed that among the 17 genes identified as leader and class B genes, 10 were previously associated with periodontitis.

The leader gene group included only five genes. A literature search confirmed that every gene we identified as a leader gene could play an important role in periodontitis at a molecular level (Table 3). For instance, NFKB1 (p50) and RELA (p65) interact to form the NFKB complex.[30] This complex triggers an intracellular pathway leading to inflammation and immune

Table 1.

Leader Genes and Class B Genes in Human Periodontitis With Number of Citations in PubMed, Number of Interactions, and Specificity Score

	Official Name	Citations (n)	Interactions in Periodontitis Gene Set (n)	Interactions in Whole STRING Database (n)	Specificity Score (%)
Leader gene					
NFKB1	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	0	11	131	8
CBL	Cas-Br-M (murine) ecotropic retroviral transforming sequence	0	10	93	11
GRB2	Growth factor receptor-bound protein 2	0	9	73	12
PIK3R1	Phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	1	9	99	9
RELA	V-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor of kappa light polypeptide gene enhancer in B-cells 3, p65 (avian)	2	8	113	7
Class B gene					
IKBKB	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta	1	8	38	21
SRC	V-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	5	8	124	6
IL6R	Interleukin-6 receptor	2	8	33	24
TRAF2	TNF receptor–associated factor 2	1	7	73	9
TNF	Tumor necrosis factor (TNF superfamily, member 2; TNFSF2)	242	7	25	28
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide	0	7	35	20
IL6ST	Interleukin-6 signal transducer (gp130, oncostatin M receptor)	2	6	33	18
IL4R	Interleukin-4 receptor	5	7	45	15
IL6	Interleukin-6 (interferon, beta 2)	2	6	33	18
TNFRSF1A	Tumor necrosis factor receptor superfamily, member 1A	0	6	50	12
CRK	V-crk sarcoma virus CT10 oncogene homolog (avian)	0	6	47	13
IL1R1	Interleukin-1 receptor, type 1	0	6	27	22

processes. In addition, a recent analysis[31] of the localization of the NFKB complex demonstrated significantly increased activity beneath periodontal lesions. Furthermore, elevated NFKB activity was demonstrated in the pathogenesis of several chronic inflammatory diseases associated with periodontitis, including atherosclerosis and type 2 diabetes. However, the two single subunits NFKB1 and RELA are actually encoded by different genes, which interact differently with the other genes involved in human periodontitis (Fig. 2). Specific roles of NFKB1 and RELA in the pathogenesis of periodontitis have not been fully elucidated. In particular, although RELA's role in periodontitis was identified recently,[30] a detailed characterization of the precise role of NFKB1 in periodontitis is lacking. A recent study[31] suggested that NFKB inhibition might be useful in treating aggressive periodontitis. Our analyses showed that NFKB1 and RELA had a very high number of interactions among identified genes involved in periodontitis. Therefore, a central role for these two genes, whose regulation is coordinated, in the development of periodontitis may be assumed, and a targeted experimentation will allow a deeper investigation into the molecular pathogenetic mechanisms involving these two genes.

PIK3R1 was identified as a genomic marker for severe periodontitis in a large-scale epidemiologic study by Suzuki et al.[32] This study focused on the identification of single nucleotide polymorphisms (SNPs) in 244 genes related to systemic and severe

Table 2.

Genes Sorted According to Number of Citations Among the 61 Genes Identified to Be Specifically Involved in Human Periodontitis

Gene Name	Official Name	Citations (n)	Interactions (n)	Class
IL1A	Interleukin-1, alpha	403	5	C
IL1B	Interleukin-1, beta	403	5	C
TNF	Tumor necrosis factor (TNF superfamily, member 2; TNFSF2)	202	7	B
CD14	CD14 antigen	54	1	I
TNFSF11	Tumor necrosis factor (ligand) superfamily, member 11	37	2	H
IL10	Interleukin-10	36	3	E
MMP1	Matrix metallopeptidase 1 (interstitial collagenase)	33	0	Orphan
MMP2	Matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase)	29	3	F
TLR4	Toll-like receptor 4	25	1	I
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	22	2	H

Table 3.

Established or Putative Role of Leader Genes in Human Periodontitis

NFKB1	Increased activity beneath periodontal lesions; elevated activity in the pathogenesis of several chronic inflammatory diseases associated with periodontitis (atherosclerosis and diabetes)
CBL	Possible involvement of this gene in bone resorption
GRB2	Possible involvement in control of epithelial growth and differentiation in periodontal tissues
PIK3R1	Important marker of severe periodontitis
RELA	Increased activity beneath periodontal lesions

periodontitis. SNPs are DNA sequence variations that result from the alteration of a single nucleotide, which may cause altered gene function and/or altered gene expression. It was observed that some SNPs in PIK3R1 were significantly more common in subjects with severe periodontitis than in healthy volunteers. Our study identified PIK3R1 as a leader gene: this may be further strong confirmation of its importance in the development of periodontitis. However, a targeted experimentation is required to fully elucidate its role in this disease.

The other two leader genes identified, CBL and GRB2, have not been described in association with periodontitis. CBL is an adaptor protein for receptor protein-tyrosine kinases and positively regulates receptor protein-tyrosine kinase ubiquitination. Ubiquitination of receptor protein-tyrosine kinases terminates signaling by marking the active receptors for degradation. Bruzzaniti et al.[33] found that CBL participates in signaling complexes that are important in the assembly and remodeling of the actin cytoskeleton, leading to changes in osteoclast adhesion, migration, and resorption. Moreover, it was shown that CBL affects migration and bone-resorbing activity in osteoclasts in a vitro model.[34] Therefore, it may be interesting to set up an ad hoc experiment to evaluate the involvement of this gene in bone resorption during acute phases of periodontitis.

GRB2 is an essential component of epidermal growth factor receptor signaling to retrovirus associated sequences, which was described to be involved in the control of epithelial growth and differentiation in periodontal tissues.[35] Because CBL and GRB2 have a very high number of interactions and are closely linked to PIK3R1 (Fig. 2), it can be assumed that these two genes may be important in the pathogenesis of periodontitis. A targeted experimentation, currently in progress, may shed new light on their role in this disease.

With regard to class B genes, the literature search revealed that nine of 12 genes have some evidence of involvement in periodontitis. In particular, some class B genes (IL6, IL6ST, IL1R1, IL6R, and IL4R) belong to the interleukin or interleukin receptor classes. These classes of proteins play an important role in the development of severe periodontitis.[36,37] For instance, a polymorphism in the IL6R gene was significantly associated with aggressive and chronic periodontitis in a recent study[38] conducted on >400 subjects. Similar results were found for IL6 and IL6ST.[39] IKBKB encodes for a kinase directly associated with NFKB complex activation. Results from a biochemical study[40] suggested the central role of the NFKB complex activation in the early development of periodontitis and proposed it as a potential therapeutic target. TNF involvement in periodontitis is largely established.[41] Present research in periodontitis is focusing on this gene (>40 articles were published regarding TNF and periodontitis in 2006). Our *ab initio* analysis seemed to confirm the important role of TNF in periodontitis, due, at least in part, to its high number of interactions with the other genes involved in this disease. The role of soluble TNF receptor (TRAF2) in periodontitis was suggested by a biochemical study,[42] which proposed that an imbalance in its regulation in gingival crevicular fluid might be related to chronic periodontitis severity. SRC, another identified class B gene, is known to be involved in immunoregulation mechanisms in oral mucosa during chronic periodontitis.[43] CRK has not been described in association with periodontitis. This gene encodes for SH2-SH3 adaptor protein. Our analysis showed that this gene is directly linked with three of five proposed leader genes (PIK3R1, GRB2, and CBL); therefore, it may be assumed to have an important role in periodontitis development. With regard to the other two class B genes (TNFRSF1A and PDGFRB), several direct interactions with other leader and class B genes are established; their potential role in periodontitis may be assumed and investigated further with targeted experimentations. However, to limit at least partially database-related bias, specificity of leader and class B genes to human periodontitis should be verified.

Specificity of Identified Genes to Human Periodontitis

Analysis of specificity score revealed a lower mean specificity for leader genes than for class B genes. Specificity score was defined herein to determine whether the high interaction scores between the leader genes and class B genes was related to their belonging to the given gene set (i.e., periodontitis related) or is caused simply by the fact that the given gene has an overall high interaction score over the entire given database (i.e., STRING). Very low specificity scores, e.g., <1%, most likely reflect the latter rather than the former alternative and indicate that the results are likely a consequence of database bias. When combined with the partly recursive nature of the used algorithm of gene identification, which includes the initial choice of the key words, the subsequent STRING-based expansion, and finally cross-checking with a subset of the initial key words, the database bias can be a source of circular reasoning that might affect the validity of the results. In this study, the specificity scores averaged 9.4% for leader genes and 17.2% for class B genes, which suggests that neither is caused by database bias, and we can consider that circular reasoning has been at least partially ruled out. However, it is important that class B genes had markedly higher specificity than leader genes. A possible explanation is that class B genes characterize the given disorder by itself. It also might be possible that the leader genes are links between periodontitis and other complex diseases clinically related to periodontitis, such as diabetes or cardiac ischemia.[29] At present, we are pursuing the bioinformatic identification of such possible links.

Limitations

The results of this study should be considered more as well-supported hypotheses than as proven statements. Our *ab initio* analysis used data mining, i.e., sorting through large amounts of data and picking up relevant information to potentially discover new knowledge. Therefore, because this approach is completely based on previous information, it is only able to generate new hypotheses, which should be verified with proper experimentation, e.g., microarrays. However, the circular reasoning risk was limited, at least in part, by the calculation of specificity scores that seemed to suggest that database bias did not affect the results, which are hypothetical in nature. Therefore, a limited circular reasoning–related bias might not represent a problem, because the results will be confirmed by experimentation. Moreover, our theoretical results are well supported by literature findings on the contribution of single genes to periodontitis as follows from the above described multiple experimental evidence.

CONCLUSIONS

Our data suggested that an approach based on bioinformatic and statistical analysis of already existing databases could be a starting point to improve our knowledge about periodontitis, as well as other complex diseases, and to plan targeted experimentation. However, previous studies[20,44] showed that only the combination of experimental and theoretical results can provide a significant picture of a complex phenomenon, such as gene expression in a particular biologic system. With the results of the present *ab initio*

study, we identified some genes with a potential important role in periodontitis, some of which have never been associated with this pathology. On this basis, we are planning targeted experimentations that should shed new light on the molecular mechanisms of periodontitis.

Most of our completely *ab initio*–identified leader and class B genes have a proven epidemiologic, clinical, or biochemical involvement in periodontitis. The ones with no proven evidence are directly linked to other leader or class B genes with a definite role in periodontitis development and, thus, may be assumed to have a potential role in such disease, which may be elucidated by future targeted studies.

Targeted studies will allow the elucidation of the role of these genes in periodontitis and a further analysis of their correlations with other genes, as we theoretically suggest in this work. The 10 genes with the largest number of citations in this pathology showed a low importance in terms of the number of interactions. This is further confirmation of the need for new targeted molecular experimentations in periodontitis, for instance by DNA microarrays, based on the genes with the highest importance in this pathology. As suggested by the present and previous studies, bioinformatics can give added value to molecular biology. In particular, the detailed analysis of gene interaction maps and the ranking of genes according to their importance might have great value in the identification of new targets for a focused experimental analysis, which may confirm each hypothesis and suggest potential risk factors and therapy targets. In this theoretical study, we identified genes involved or potentially involved in periodontitis, and we identified the genes with the highest number of interactions in this pathology, i.e., the leader genes. For some of these genes (GRB2, CBL, and NFKB1), a detailed description of their role in periodontitis is still lacking, whereas the other leader genes (RELA and PIK3R1) have not been fully characterized. At the same time, orphan genes merit further characterization. Therefore, these findings can suggest targeted experimental analysis. A possible way to provide further details about molecular mechanisms of periodontitis is the use of targeted DNA and protein microarrays, displaying only selected genes. In this way, it may be possible to confirm and expand bioinformatics-based predictions of molecular mechanisms underlying periodontitis.

ACKNOWLEDGMENTS

This project was supported by grants to the Electronics Biotechnology Advanced Foundation, Rome, Italy, by the Italian Ministry of Research and University (MIUR) for "Operation" and by a Foundation for Investments in Basic Research, Ministry of Instruction, Rome, Italy, International Grant on Functional Proteomics and Cell Cycle Progression (RBIN04RXHS) from MIUR to Interuniversities Research and Didactic Service Center on BioOrganic Nanosciences and Nanotechnologies, Nanoworld Institute of the University of Genoa. The authors report no conflicts of interest related to this study.

REFERENCES

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931-945.
2. Haffajee AD, Uzel NG, Arguello EI, Torresyap G, Guerrero DM, Socransky SS. Clinical and microbiological changes associated with the use of combined antimicrobial therapies to treat "refractory" periodontitis. *J Clin Periodontol* 2004;31:869-877.
3. Hirschfeld L, Wasserman B. A long-term survey of tooth loss in 600 treated periodontal patients. *J Periodontol* 1978;49:225-237.
4. McFall WT Jr. Tooth loss in 100 treated patients with periodontal disease: A long-term study. *J Periodontol* 1982;53:539-549.
5. Armitage GC. Development of a classification system for periodontal diseases and conditions. *Ann Periodontol* 1999;4:1-6.
6. Michalowicz BS, Diehl SR, Gunsolley JC, et al. Evidence of a substantial genetic basis for risk of adult periodontitis. *J Periodontol* 2000;71:1699-1707.
7. Hart TC. Genetic considerations of risk in human periodontal disease. *Curr Opin Periodontol* 1994;3-11.
8. Walker SJ, Van Dyke TE, Rich S, Kornman KS, di Giovine FS, Hart TC. Genetic polymorphisms of the IL-1alpha and IL-1beta genes in African-American LJP patients and an African-American control population. *J Periodontol* 2000;71:723-728.
9. Loos BG, John RP, Laine ML. Identification of genetic risk factors for periodontitis and possible mechanisms of action. *J Clin Periodontol* 2005;32(Suppl. 6):159-179.
10. Tabor HK, Risch NJ, Myers RM. Opinion: Candidate-gene approaches for studying complex genetic traits: Practical considerations. *Nat Rev Genet* 2002;3:391-397.
11. Hirschhorn JN. Genetic approaches to studying common diseases and complex traits. *Pediatr Res* 2005; 57:74R-77R.
12. von Mering C, Jensen LJ, Snel B, et al. STRING: Known and predicted protein-protein associations, integrated and transferred across organism. *Nucleic Acids Res* 2005;33:D433-D437.
13. Ashton GH, McGrath JA, Sonth AP. Strategies to identify disease genes. *Drugs Today* 2002;38:235-238.
14. Kim DM, Ramoni MF, Nevins M, Fiorellini JP. The gene expression profile in refractory periodontitis patients. *J Periodontol* 2006;77:1043-1050.
15. Steinberg T, Dannewitz B, Tomakidi P, et al. Analysis of interleukin-1beta-modulated mRNA gene transcription in human gingival keratinocytes by epithelia-specific cDNA microarrays. *J Periodontal Res* 2006;41:426-446.
16. Ramachandran N, Hainsworth E, Bhullar B, et al. Self-assembling protein microarrays. *Science* 2004;305: 86-90.
17. Giacomelli L, Nicolini C. Gene expression of human T lymphocytes cell cycle: Experimental and bioinformatic analysis. *J Cell Biochem* 2006;99:1326-1333.
18. Sivozhelezov V, Giacomelli L, Triphati S, Nicolini C. Gene expression in the cell cycle of human T

lymphocytes: I. Predicted gene and protein networks. *J Cell Biochem* 2006;97:1137-1150.
19. Nicolini C, Spera R, Stura E, Fiordoro S, Giacomelli L. Gene expression in the cell cycle of human T-lymphocytes: II. Experimental determination by DNASER technology. *J Cell Biochem* 2006;97:1151-1159.
20. Nicolini C. Nanogenomics for medicine. *Nanomedicine* 2006;1:147-151.
21. Nicolini C, Malvezzi M, Tomaselli A, Sposito D, Tropiano G, Borgogno E. DNASER I: Layout and data analysis. *IEEE Trans Nanobioscience* 2002;1:67-72.
22. Sivozhelezov V, Braud C, Giacomelli L, et al. Immunosuppressive drug-free operational immune tolerance in human kidney transplants recipients: Part II. Non-statistical gene microarray analysis. *J Cell Biochem* 2008;103:1693-1706.
23. Gurkan C, Lapp H, Hogenesch JB, Balch WE. Exploring trafficking GTPase function by mRNA expression profiling: Use of the SymAtlas web-application and the Membrome datasets. *Methods Enzymol* 2005;403:1-10.
24. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: Integrating information about genes, proteins and diseases. *Trends Genet* 1997;13:163.
25. Tsai CA, Lee TC, Ho IC, Yang UC, Chen CH, Chen JJ. Multi-class clustering and prediction in the analysis of microarray data. *Math Biosci* 2005;193:79-100.
26. Shannon W, Culverhouse R, Duncan J. Analyzing microarray data using cluster analysis. *Pharmacogenomics* 2003;4:41-52.
27. Kinane DF, Shiba H, Hart TC. The genetic basis of periodontitis. *Periodontol 2000* 2005;39:91-117.
28. Kinane DF, Lappin DF. Immune processes in periodontal disease: A review. *Ann Periodontol* 2002;7:62-71.
29. Amar S, Han X. The impact of periodontal infection on systemic diseases. *Med Sci Monit* 2003;9:RA291-RA299.
30. Tipton DA, Gay DC, DeCoster VA. Effect of a cyclooxygenase-2 inhibitor on interleukin-1b-stimulated activation of the transcription factor nuclear factor-kappa B in human gingival fibroblasts. *J Periodontol* 2007;78:542-549.
31. Milward MR, Chapple IL, Wright HJ, Millard JL, Matthews JB, Cooper PR. Differential activation of NF-kappaB and gene expression in oral epithelial cells by periodontal pathogens. *Clin Exp Immunol* 2007;148:307-324.
32. Suzuki A, Ji G, Numabe Y, Ishii K, Muramatsu M, Kamoi K. Large-scale investigation of genomic markers for severe periodontitis. *Odontology* 2004;92:43-47.
33. Bruzzaniti A, Neff L, Sanjay A, Horne WC, De Camilli P, Baron R. Dynamin forms a Src kinase-sensitive complex with Cbl and regulates podosomes and osteoclast activity. *Mol Biol Cell* 2005;16:3301-3313.
34. Horne WC, Sanjay A, Bruzzaniti A, Baron R. The role(s) of Src kinase and Cbl proteins in the regulation of osteoclast differentiation and function. *Immunol Rev* 2005;208:106-125.
35. Gao Q, Zhang S, Jian X, Zeng Q, Ren L. Expression of epidermal growth factor and epidermal growth factor receptor in rat periodontal tissues during orthodontic tooth movement. *Zhonghua Kou Qiang Yi Xue Za Zhi* 2002;37:294-296.
36. Pietruska M, Zak J, Wysocka J, et al. Evaluation of selected peripheral blood leukocyte functions in patients with various forms of periodontal disease. *Arch Immunol Ther Exp (Warsz)* 2004;52:208-212.
37. Lin TW, Cardenas L, Soslowosky LJ. Tendon properties in interleukin-4 and interleukin-6 knockout mice. *J Biomech* 2005;38:99-105.
38. Galicia JC, Tai H, Komatsu Y, Shimada Y, Ikezawa I, Yoshie H. Interleukin-6 receptor gene polymorphisms and periodontitis in a non-smoking Japanese population. *J Clin Periodontol* 2006;33:704-709.
39. Suzuki A, Ji G, Numabe Y, et al. Single nucleotide polymorphisms associated with aggressive periodontitis and severe chronic periodontitis in Japanese. *Biochem Biophys Res Commun* 2004;317:887-892.
40. Carayol N, Chen Y, Yang F, et al. A dominant function of IKK/NF-kappaB signaling in global lipopolysaccharide-induced gene expression. *J Biol Chem* 2006;281:31142-31151.
41. Beklen A, Laine M, Venta I, Hyrzkas T, Konttinen YT. Role of TNF-alpha and its receptors in pericoronitis. *J Dent Res* 2005;84:1178-1182.
42. Ikezawa I, Tai H, Shimada Y, Komatsu Y, Galicia JC, Yoshie H. Imbalance between soluble tumour necrosis factor receptors type 1 and 2 in chronic periodontitis. *J Clin Periodontol* 2005;32:1047-1054.
43. Muthukuru M, Cutler CW. Upregulation of immunoregulatory Src homology 2 molecule containing inositol phosphatase and mononuclear cell hyporesponsiveness in oral mucosa during chronic periodontitis. *Infect Immun* 2006;74:1431-1435.
44. Braud C, Baeten D, Giral M, Soulillou JP, Brouard S. Immunosuppressive drug-free operational immune tolerance in human kidney transplant recipient: I. Blood gene expression statistical analysis. *J Cell Biochem* 2008;103:1681-1692.

Correspondence: Prof. Claudio Nicolini, Interuniversities Research and Didactic Service Center on BioOrganic Nanosciences and Nanotechnologies, University of Genoa and Nanoworld Institute, Corso Europa 30, 16132 Genoa, Italy. Fax: 39-010-35338215; e-mail: manuscript@ibf.unige.it.

Submitted January 30, 2008; accepted for publication April 7, 2008.