Gene Expression Meta-Analysis Reveals Gene Associated with Periodontitis

Mingi Kim *mingikim1628@gmail.com*
Lumiere Education Inc.
Portland, 97229, USA

Laura K. Harris *oesterei@gmail.com*
Harris Interdisciplinary Education
Dewitt, 48820, USA
Lumiere Education Inc.
Wilmington, 48824, USA

Abstract

Background: Periodontitis is a severe form of gum disease which in flames and destroys tissues supporting the teeth and leading to potential loss of teeth. If untreated, it could lead to heightened risk of serious health issues and other complications. This study was performed to meta-analyze gene expression data from Periodontitis infected samples to possibly identify clinically targetable genes associated with Periodontitis.

Methods: This study defined 8 gene signatures between Periodontitis infected and mock infected from 2 datasets and compared them through Gene Set Enrichment Analysis (GSEA). Positive and negative panels were defined from GSEA identified leading-edge genes in each query set. GSEA was then utilized to calculate enrichment and identify differentiation in gene expression of Periodontitis and compared it with other gene signatures.

Results: GSEA identified 93 and 127 probes, representing 71 and 114 unique positive and negative panels, respectively, from the GSE10334 dataset. Also, GSEA identified 110 and 112 probes, representing 92 and 108 unique positive and negative panels, respectively, from the GSE16134 dataset. Non-random significant enrichment was observed between each identified panel and all verification gene signatures derived from both datasets and 23 overexpressed and 33 under expressed genes were found out to be shared across all verification gene signatures. The final genes identified to have significant association with Periodontitis consisted of few genes that were found to be associated with periodontitis in previous studies and also showed new genes that have not been identified to have connection with periodontitis previously.

Conclusion: Gene Set Enrichment Analysis identified genes associated with Periodontitis with prior or no prior known connection with Periodontitis. These results might be useful as potential therapeutic targets for Periodontitis disease.

Keywords: Periodontitis, Microarray, Gene expression, Meta-Analysis, Gene Set Enrichment Analysis.

1. INTRODUCTION

Periodontitis is a common disease in the oral cavity, and it is to have effect on about 33% of individuals in the United States (Kwon et al., 2021). Gingivitis denotes the early stage of periodontitis, displaying a relatively less severe clinical condition, and is widely accepted as a condition with a favorable treatment prognosis (Trombelli et al., 2018). If Gingivitis is not addressed for a period of time, then it will make its progression into periodontitis (Abusleme et al., 2021). Few symptoms and risk of periodontitis disease are gingival tissue inflammation, alveolar

bone loss, and reduction or transformation of periodontal ligament (Tonetti et al., 2018). Although periodontitis can be caused by multiple different contributing factors, the main cause of the disease is formation of bacterial biofilm on the tooth surfaces (Hajishengallis et al., 2017). Risk factors that can affect periodontitis are environment, genetic, medication, systemic health, and lifestyle of the patient, which may determine the progress and severity of the disease (Jepsen et al., 2018). If periodontitis is left untreated, it can cause tooth loss, issues with chewing functions, and array of social and medical afflictions (Sanz et al., 2020). There is no full cure to periodontitis because it is caused by a bacterial biofilm that is formed in a person's mouth daily; however, periodontitis can be managed by mechanical removal and other periodontal therapy (Sanz et al., 2020). In periodontal therapy, the effectiveness of its treatment can be influenced by multiple factors such as the patient's environment or response may affect the treatment progress of the disease (Sanz et al., 2020). Identifying the genetic factors related to periodontitis is important information to find new treatment plans for patients, or possibly even the cure for the disease.

The development and progression of periodontitis are influenced by the dynamic interactions between the bacterial biofilm and the immune system (Kim et al., 2016). Other factors such as genetics, environment, medication, systemic health, and lifestyle of the patient were proven to have effect on pathogenesis and progression of the disease (Li et al., 2014)(Lundmark et al., 2018). DNA, or deoxyribonucleic acid, is a double-stranded macromolecule composed of nucleotides that carries the genetic instructions necessary for the development, functioning, and reproduction of all known living organisms (Kukurba et al., 2015). RNA acts as an intermediary to transcribe the genetic information encoded in DNA into functional proteins (Kukurba et al., 2015). The utilization of RNA sequencing as the primary data collection method in this experiment allows for the detection and analysis of gene expression levels, enabling insights into cellular behavior (Kim et al., 2016). The key benefits of RNA sequencing are being able to directly sequence RNA enabling to take an unbiased approach toward genome sequence and RNA sequencing exhibits exceptional accuracy in detecting gene expression due to its wide dynamic detection range and minimal background noise (Kim et al., 2016). RNA sequencing first starts by extracting RNA from target cells(Peymani et al., 2022)(Sczepanik et al. 2020). RNA will be then reverse transcribed into Complementary DNA (cDNA) fragments with adapters, which contains essential functions to enable sequencing, added on to each end. Then the genes were sequenced into the NGS platform(Peymani et al., 2022)(Sczepanik et al. 2020).

Conducting a comprehensive exploration of gene expression profiles linked to periodontitis is crucial in order to find out a possible genome that could be targeted for treatment of the disease (Jeon et al., 2020)(Gao et al., 2024)(Cho et al., 2021). There have been many studies aimed at understanding the molecular changes associated with periodontitis by examining differential gene expression through experimental techniques such as robust multi-array analysis or model-based clustering (Demmer et al., 2008)(Kebschull et al., 2014). This approach generated a list of most differentially expressed genes that met statistical criteria to be considered biologically relevant to periodontitis. Additional investigation is required to fine-tune lists of identified genes and pinpoint the most promising gene candidates for therapeutic targeting.

Various computational methods have been employed to analyze experimentally derived lists of genes that exhibit differential expression such as multiarray analysis and model-based clustering discussed in previous paragraph (Papapanou et al., 2018). For example, log scale robust multi-array analysis (RMA) was employed for the preprocessing of gene expression data from microarray experiments (Wang et al., 2023). RMA involves the normalization and summarization of expression data to obtain reliable measurements. The subsequent differential expression analysis utilized a mixed-effects linear model, considering patient effects as random variables and gingival tissue status as a fixed effect with two levels ("healthy" vs. "diseased"). Statistical significance was assessed using both the Bonferroni criterion and q-value, and fold-change values are calculated to represent the relative RNA levels between "diseased" and "healthy" samples for each probe set. This study identified several positively and negatively differentially expressed genes in Periodontitis infected samples. Gene Set Enrichment Analysis (GSEA) utilizes established gene sets from the public knowledge base in gene signature to consider all

gene in expression dataset (Park et al., 2021). There are unique advantages of using Gene Set Enrichment Analysis (GSEA) to analyze genes (Park et al., 2021). For example, it is able toconsider the whole gene dataset, so it does not exclude any genes for not meeting the cutoff. Further, Gene Set Enrichment Analysis (GSEA) can be used to confirm and verify enrichment for identified query gene sets defined (Park et al., 2021). Comparisons between multi-array analysis and Gene Set Enrichment Analysis (GSEA) have not been made previously, so this experiment will use a GSEA-based method to identify what similarities and differences in the resulting dataset would be found.

There remain numerous unknown details of genes related to periodontitis. Stronger conclusive evidence is imperative to establish the true responsibility of previously pinpointed genes in causing periodontitis. This study contributes to bridging these gaps within the broader scientific literature, serving as the primary investigation utilizing microarray technology for analysis. The earlier application of a GSEA-based analysis method in related studies facilitated the identification of genes associated with severe acute respiratory syndrome infection by comparing differential gene expression with mRNA expression datasets (Park et al., 2021). This study will incorporate Gene Set Enrichment Analysis (GSEA) by examining gene expression data to calculate differential expressions of genes to identify specific genes that have strong relationships with Periodontitis. This will provide possible future insight to develop a possible cure by targeting genes related to Periodontitis.

2. LITERATURE REVIEW

2.1 Gene Expression Datasets Used

Gene Expression Omnibus (GEO) was utilized to gather datasets for use in the study. The publication status, number and type of control and periodontitis samples, and platform used to profile samples were considered when selecting datasets. Based on this criteria, 2 published datasets - GSE10334 and GSE16134 - were selected for their large sample size and shared platform. Briefly, GSE10334 analyzed a total of 247 gingival tissue samples - 183 diseased and 64 healthy interproximal papilla - from 90 non-smokers with varying degree of periodontitis - 63 chronic and 27 aggressive (PMID: 18980520). GSE16134 analyzed 310 gingival tissue samples - 241 diseased and 69 unaffected interproximal papilla - from 120 systematically healthy patients undergoing periodontal surgery (PMID: 19835625, PMID: 24646639). Both datasets profiled samples on 54675 probes from the Affymetrix HU-U133Plus 2.0 array (GEO GPL570). Analyzing datasets with the same platform is advantageous for preserving data integrity, as it avoids the need for converting platform probe identifiers and potential duplication of gene identifiers that could result in data loss during GSEA.

This study used an inductive research approach to analyze gene expression data related to periodontitis. The research design was structured as an exploratory investigation to discover novel biological insights from the data. Gene expression data on periodontitis was collected from Gene Expression Omnibus (GEO) database to allow meta-analysis between datasets. We took an inductive research approach by utilizing Gene Set Enrichment Analysis (GSEA) to analyze the complete gene expression. We aimed to discover broader pathway-level molecular signatures associated with periodontitis. The gene expression data was ranked based on t-scores calculated between periodontitis and healthy control samples. Then distribution of gene sets was analyzed within a ranked dataset to identify pathways exhibiting significant enrichment. The results were analyzed to pinpoint enriched pathways that may be implicated in periodontitis pathogenesis.

2.2 Signature Formation

Data from each GEO dataset was split randomly into 3 smaller datasets to maximize the utility of these datasets (Table 1). An attempt to have equal numbers of periodontitis and control samples per subset was made, but was unsuccessful due to the number of total samples. Therefore, the 2 subsets from each dataset that had an identical number of samples were used for identification purposes and the remaining subset used for verification of identification results. Further, each

dataset was used in its entirety for additional independent verification. This generated 8 groups of data for subsequent analysis.

Dataset	Group 1 (N)	Group 2 (N)	Group Name	Use
GSE10334	Periodontitis (61)	Control (21)	PeriodovsContvID1	ID
	Periodontitis (61)	Control (21)	PeriodovsContvID2	ID
	Periodontitis (61)	Control (22)	PeriodovsContvV	V
	Periodontitis (241)	Control (69)	PeriodovsContvall	V
GSE16134	Periodontitis (80)	Control (23)	PeriodovsContvID1	ID
	Periodontitis (80)	Control (23)	PeriodovsContvID2	ID
	Periodontitis (81)	Control (23)	PeriodovsContvV	V
	Periodontitis (183)	Control (64)	PeriodovsContvall	V

TABLE 1: Data groups generated for use in this study (ID, identification; N, number of samples; V, verification).

In the pursuit of reliable and reproducible results, it is crucial to ensure the consistency of gene expression data. To achieve this, each dataset and its subsequent subsets were normalized separately using z-score. This normalization method not only enhanced the comparability of gene expression levels within and across datasets but also minimized the impact of potential outliers, providing a robust foundation for subsequent analyses and interpretations.

Signatures - lists of probes ranked by differential expression between periodontitis and control samples using T-score - were generated for each dataset and its subsequent subsets (Table 2). Briefly, probes that were over-expressed in periodontitis compared to control samples have a positive T-score and would fall within the positive tail of the signature. Similarly, probes that were under-expressed in periodontitis compared to control samples would fall in the negative tail of signature from their negative T-scores, respectively. probes that had little change in expression between periodontitis and control samples fall within the middle of gene signature with a T-score around zero and is reflected in where the signature crosses from a positive to a negative T-score. For additional normalization and to avoid signature skewing, T-scores in each signature were adjusted by the T-score of the probe at the signature's mid-point so that probes in the signature were balanced between positive and negative T-scores.

Dataset	Signature	Use	High	Low	Cross	Adjustment
GSE10334	PeriodovsContvID1	ID	12.7	-16.2	28795	-0.34
	PeriodovsContvID2	ID	10.4	-8.7	27493	-0.02
	PeriodovsContvV	V	12.3	-10.9	24984	0.39
	PeriodovsContvall	V	13.2	-13.6	28946	0.17
GSE16134	PeriodovsContvID1	ID	13.0	-13.0	25778	0.19
	PeriodovsContvID2	ID	9.4	-10.6	30533	-0.37

	PeriodovsContvV	V	15.9	-13.5	24820	0.46
	PeriodovsContvall	V	12.3	-10.9	24984	-0.24

TABLE 2: Signatures generated for use in this study (ID, identification; N, number of samples; V, verification).

2.3 Gene Set Enrichment Analysis Based Meta-analysis

First, probes associated with periodontitis needed to be identified. To begin this process, the 500 most positive and negative differentially expressed probes were selected from each identification signature to generate 4 query gene sets per dataset. The number 500 was chosen to capture the maximum coverage of signatures allowed by GSEA (PMID: 16199517, PMID: 34367154, https://www.cscjournals.org/library/manuscriptinfo.php?mc=IJBB-263), which was subsequently used to compare each identification signature to the 4 query sets from its associated dataset. Briefly, GSEA used each gene's T-score from the reference signature to calculate a running enrichment score between an individual query set and reference signature. Probes that contributed to reaching the maximum enrichment score were selected as leading-edge genes. To establish a null distribution and estimate the significance of that achieved maximum enrichment score, GSEA ran 1000 reference signature permutations to get a normalized enrichment score. In this way GSEA was used as an estimating method to evaluate the enrichment value of query sets compared to the reference signature. Leading-edge probes shared between identification signatures were collected to form positive and negative periodontitis panels for each dataset. For completeness, pathway enrichment analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 was performed on each panel to understand biological meanings behind these list of gene symbols representing probes (PMID: 19033363, PMID: 19131956).

Probes associated with periodontitis needed to be verified. To do this, periodontitis panels were compared to verification signatures using GSEA. To assess if results generated from GSEA could be achieved randomly, 1000 randomly selected panels the average size between positive and negative periodontitis panels from the same dataset were used as query sets against the dataset's verification signatures to generate a null distribution of NES. From this, a distribution p-value was calculated by comparing the NES achieved by positive and negative periodontitis panels to the null distribution of NES and counting the number of equal or better NES to estimate significance. Further, resulting leading-edge genes from these analyses were examined for similarities, and probes found across all verification leading-edges were selected for heat map visualization across all signatures. All python scripts used for this work are freely available at GitHub located at https://github.com/oesterei/JNBs4periodontitisIJBB.

3. RESULTS

3.1 Gene Set Enrichment Analysis Based Meta-analysis Identified Periodontitis Probes

The first step in this study was to identify genes associated with periodontitis. To do this, tails from 4 identification signatures - GSE10334 PeriodovsContvID1, GSE10334 PeriodovsContvID2, GSE16134 PeriodovsContvID1, and GSE16134 PeriodovsContvID2 - were used to define 8 query gene sets (unadjusted T-score>5.37 adjusted, T-score>5.35 and unadjusted T-score<-5.75, adjusted T-score<-5.77 for positive and negative tails, respectively). Probes included in these query sets and their details were provided in STable 1. GSEA was used to calculate enrichment between identification signatures and their respective query sets from the same dataset to assess signature similarity. All analyses achieved statistically significant enrichment as seen in Table 3.

Dataset	Reference	Query	NES
GSE10334	PeriodovsContvID1	Positive tail PeriodovsContvID2	5.25
GSE10334	PeriodovsContvID1	Negative tail PeriodovsContvID2	-4.68
GSE10334	PeriodovsContvID2	Positive tail PeriodovsContvID1	4.54
GSE10334	PeriodovsContvID2	Negative tail PeriodovsContvID1	-4.91
GSE16134	PeriodovsContvID1	Positive tail PeriodovsContvID2	4.35
GSE16134	PeriodovsContvID1	Negative tail PeriodovsContvID2	-4.13
GSE16134	PeriodovsContvID2	Positive tail PeriodovsContvID1	4.78
GSE16134	PeriodovsContvID2	Negative tail PeriodovsContvID1	-3.99

TABLE 3: Enrichment of identification signatures (All Gene Set Enrichment Analysis nominal p-value and false discovery rate q-values <0.001).

Next, leading-edge membership was examined across GSEA involving identification signatures. Probes shared between leading-edges for the same dataset were used to define positive and negative periodontitis panels for that dataset. STable 2 detailed these 4 panels that were summarized in Tables 4 and 5. Briefly, 93 and 127 probes, representing 71 and 114 unique genes, were selected for the GSE10334 positive and negative panels, respectively. For GSE16134, 110 and 112 probes, representing 92 and 108 unique genes, were selected for the positive and negative panels.

Panel	Gene symbols
Positive	ADPRH, APH1B, APLNR, AQP9, BMF, C2orf88, C4orf48, CACNG4, CALCRL, CATIP-AS1, CCDC88A, CD19, CD300A, CD300LF, CD40, CEACAM3, CKAP2*, CLEC1A, CNNM2, CSF3R*, CXCR4, FAM170B-AS1, FBLN1, FBXO16, FCRL3, FER1L4, FHAD1, FMO4, GRID1, HBD, ICAM2, IFNAR2, IFNL2, IGHG1*, IGHV3-54, IGKC*, IGLC1*, IGLV1-44*, IL24, KCNMB1, KIF21B, KLHDC7B, LINC00582, LINC00944, LOC101928894, LOC101929549, LRTOMT, MEI1, MME, NXPE3, PF4, PLA2G2D, POM121L12, PTPN7, RARB, RASD2, SAA1, SCAMP5, SELP, SLAMF6, SLC6A13, SMPDL3B*, STRA6, TAGAP, TDO2, TIMD4, VNN2, WAS, WISP1, ZFAT, and ZNF709
Negative	ABLIM1, ACTR10, AHNAK, ANK3, ANKRD29*, ANKRD6, ARAP2, ARL8B, ATMIN, ATP1B3, B3GALNT2, BAG5, BCL2L2, BICD2, BLOC1S2, BNIPL, BSPRY, BTF3, C1orf21, CBR3, CCT6A, CEBPG*, CNKSR3, COBL*, COQ2, CUL3, CWH43, CYB5A*, CYP2C18, DAAM1, DBI, DIP2B, DR1, DSTN, DYNLT3, ESRP1, FAM162A*, FAM57A, FBXO45*, GAR1, GAS2L1, GATAD2A, GPR89A, GRHL3, GSTA4, HSPA14, ING2, KIF13A, KPNA4, LGALS3, LGALSL, LOC100506248, LOC101928676, LOC283070, MALSU1, MAP2, MAST4, MCUR1, MIOS*, MORC4, MTF1, NCBP2-AS2, NDUFC1, NEBL*, NHSL1, NMT2, NSG1*, NUDT5, OCLN, OSBPL2, PBDC1, PDCD2, PELO, PIGY, POF1B, PPP2CB, PSMD12, PSMD6, PSMD7, PSMG4, PTPN3, PWP1, RAB11A, RIMS3, RPF1, RWDD4, SAP18, SERBP1, SGK223, SH3RF2, SIPA1L2*, SLC35E1, SMARCA4, SNRPE, SOD1, TACC2*, TFDP1, TRIM7, TSPAN5, TUBB2A, TXNL4A, UBE2E3, UCHL3, URI1, USP53, WBP4, WIPI2, XG, YES1, YWHAB, YWHAZ, ZFYVE21, ZNF57, and ZNRF1

TABLE 4: Positive and Negative GSE10334 Periodontitis Panels Defined in this Study (More than one probe associated with the same gene symbol found in the panel).

Panel	Gene symbols
Positive	ABI2, APH1B, ARHGEF2, BIN2, C4A, C5AR1, CCDC88A, CD19, CD300A, CD79A, CD93, CDK14, CHST2, CKAP2, CLEC1A, CMTM7, CPNE5, CSF3, CYAT1, CYP24A1, DENND3, DERL3, ELMO1, ENTPD1, FAM46C, FBXO16, FMNL3, FNDC3B, FPR1, GPSM3, GRID1, HOXB7, ICAM2, IFNAR2, IGH, IGHA1*, IGHG1*, IGHV3-54, IGHV4-31, IGKC, IGLC1, IGLJ3, IGLV1-40, IGLV2-5, IGLV3-19, IGLV1-44*, IKZF1, KCNJ8, LINC-PINT, LOC100131043, LOC100293211, LOC101928894, LOC101930405, MEI1, MERTK, mir-223, MME*, MMP14, NCF4, NCOA3, NEDD9, NXPE3, P2RX4, PECAM1, PRR5L, PTP4A3*, PTPN7, RAB20, RAB30, RASD2, RGS4, RNF144A-AS1, SAA2-SAA4, SCAMP5, SEL1L3, SELP, SELPLG, SERPINA1, SLAMF6, SLC2A3, SMPDL3B, ST3GAL4-AS1, TAGAP*, TDO2, THEMIS2, TIMD4, TNFSF12-TNFSF13, TUBB1, VNN2, WAS, WIPF1, and XBP1
Negative	ACTR10, AFG3L1P, AHNAK, AHNAK2, ANK3, ANKRD29, ANKRD35, ANKRD6, ARAP2, ARHGAP35, ARL9, BAG5, BNIPL, BSPRY, C10orf99, C9orf3, CA12, CBR3, CEBPG, COBL, COQ2, CPEB2, CUL3, CWH43, CYP2C18, DBI*, DENND2C, DR1, DYNLT3, EFS, EIF5, ENTPD5, EPN3, ESRP1, ESYT2, EXPH5, FAM162A*, FAM57A, FAM83B, FAM83C, GAS2L1, GATAD2A, GNL2, GRHL3, GSTA4, GTF3C6, HLF, HN1, HOOK2, ING2, IPPK, KIAA1671, LANCL1, LGALSL, LINC01133, MAF, MAL2, MAP2, MCUR1, MFHAS1, MID2, MIOS, MIR4657, MORC4, MPHOSPH10, NCBP2-AS2, NDUFC1, NEBL, NEFL, NEFM, NHSL1, NSG1, NUDT15, OSBPL6, PAQR5, PBDC1, PLA2G12A, PLA2R1, PLXDC2, PPP2CB, PSD4, PTK6, PTPN3, RALGAPA1, RIMS3, RORA, RP11-532F12.5, RPL35A, SAP18, SGK223, SH3GL3, SIPA1L2, SLC22A23, SMARCA4, SNRPE, SNX24, SOD1, SOWAHC, TACC2*, TMEM79, TRIM7, TUBB2A, TXNL4A, UCHL3, UTP11L, VTA1, ZNF57, and ZNRF1

TABLE 5: Positive and Negative GSE16134 Periodontitis Panels Defined in this Study (More than one probe associated with the same gene symbol found in the panel).

3.2 Pathway Enrichment Analysis on Periodontitis Panels Finds Several Increased Immune System Processes

DAVID was used to calculate enrichment between leading-edge panels and biological processes and pathways from BioCarta, Gene Ontology Biological Processes (GO-BP), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, and WikiPathways for a more complete understanding of the potential implications from results generated here (PMID: 10802651, 36866529, 10592173, 34788843, 37941138, 18651794). STable 3 detailed the results from these analyses. Briefly, 18 and 6 significantly enriched pathways were shared between GSE10334 and GSE16134 for positive and negative panels, respectively. Of the up-regulated pathways, only one - platelet-mediated interactions with vascular and circulating cells from WikiPathways - was not related to immune response. This finding was supported further when the examination focused on results from individual knowledge bases specifically, since it is important to not complicate result interpretation from overlapping pathways in different knowledge bases. To illustrate this, significant results from GO-BP were illustrated in Figures 1 and 2 and clearly show immune related processes, such as immune response, innate immune response, phagocytosis engulfment, phagocytosis recognition, classical complement activation, positive regulation of B cell activation, B cell receptor signaling, immunoglobulin mediated immune response, and defense response to bacterium shared between GSE10334 and GSE16134. This is an expected finding since periodontitis involves inflammation of periodontal tissue (Sczepanik et al. 2020).

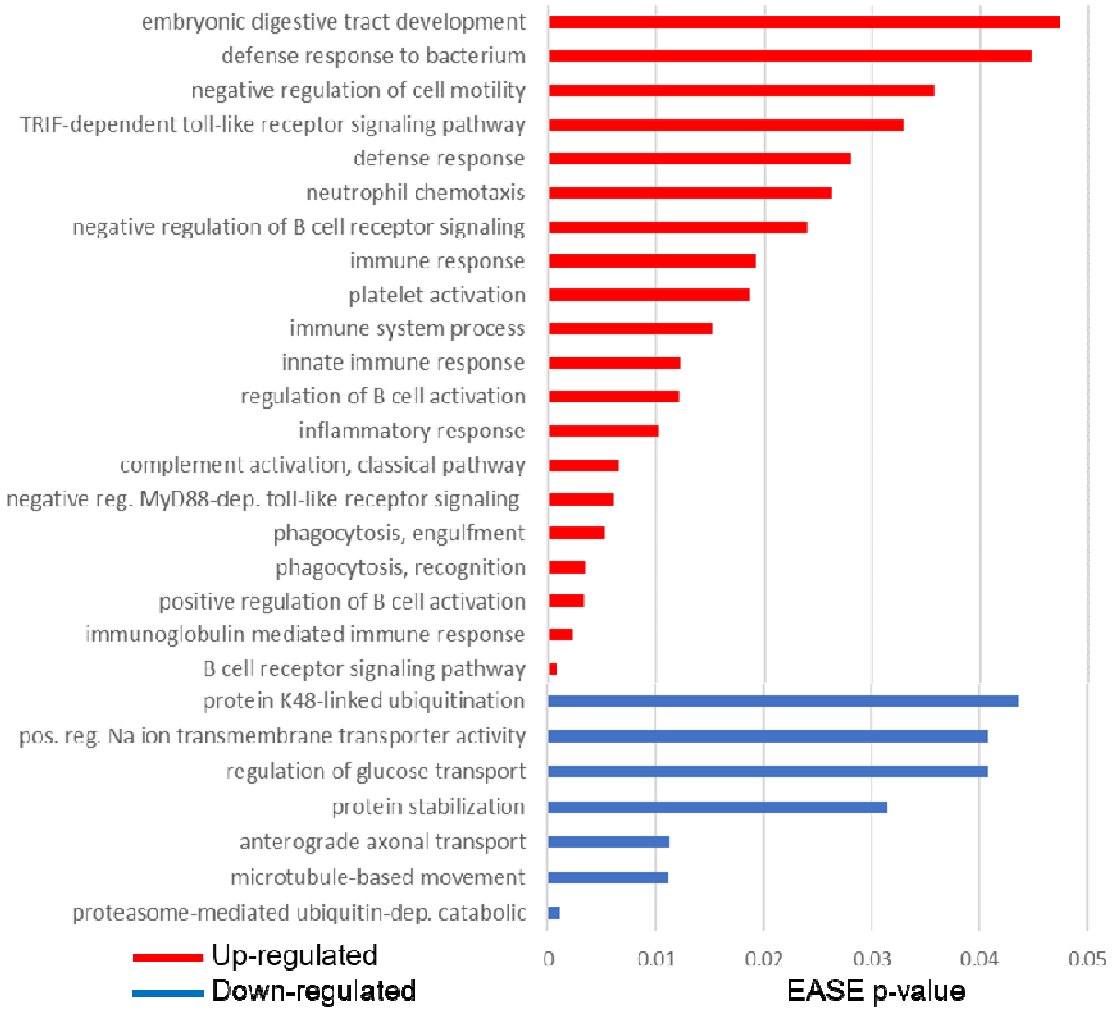

FIGURE 1: Bar Graph of EASE p-values for Significant Gene Ontology Biological Processes from GSE10334 Panels.

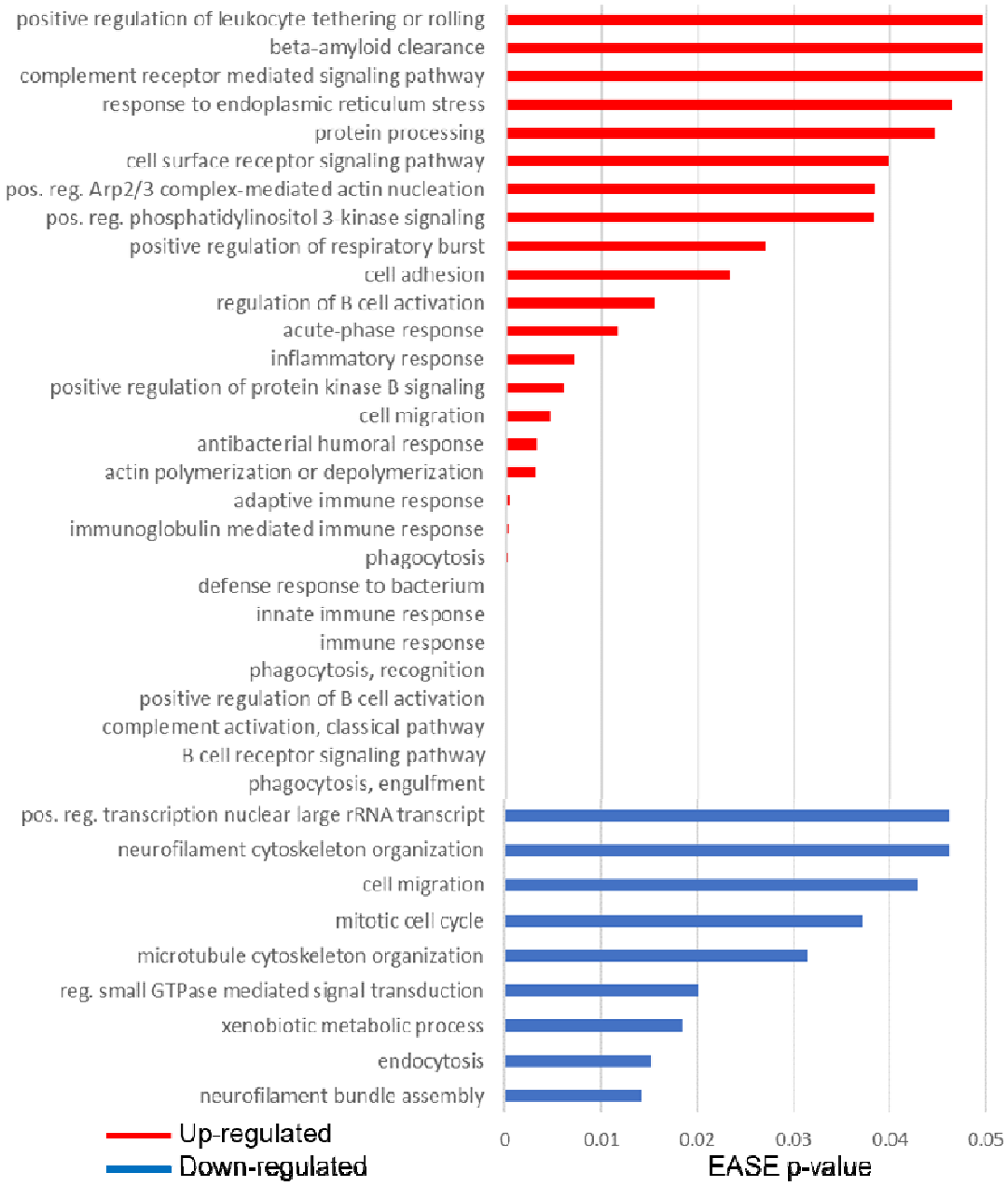

FIGURE 2: Bar Graph of EASE p-values for Significant Gene Ontology Biological Processes from GSE16134 Panels.

3.3 Periodontitis Panel Enrichment Verified in Non-overlapping and Independent Data

The next step in this study was to verify identified probes associated with periodontitis. To do this, GSEA was used to calculate enrichment of each panel compared to its associated verification signatures (Table 2). Further, random modeling was used to establish a null distribution of enrichment scores to determine the significance of the panel's enrichment. Periodontitis panels consistently achieved non-random significance in non-overlapping and independent data (Figure 3), verifying the association of panel probes with periodontitis.

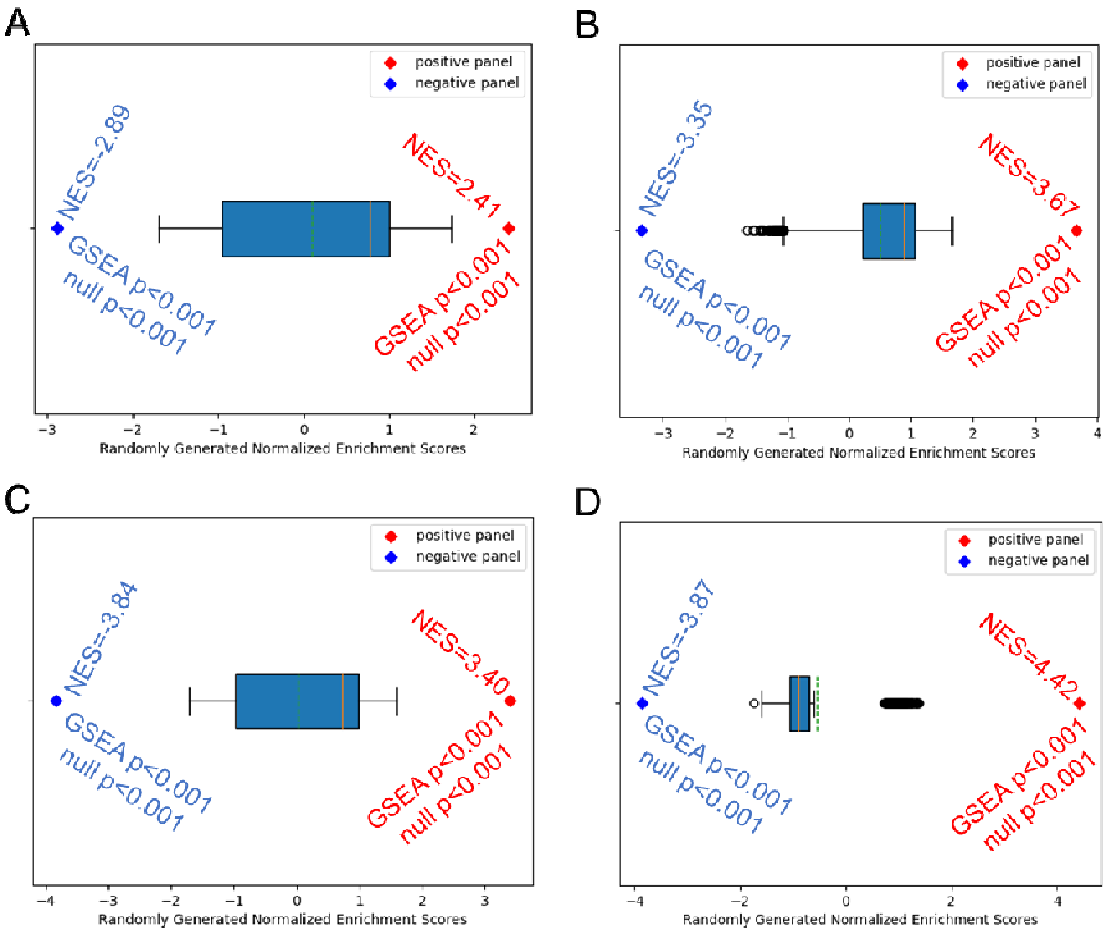

FIGURE 3: Random Modeling Showed Consistent, Non-random Enrichment of Periodontitis Panels in Verification Signatures. Random models consisted of 1000 NES from randomly selected panels (query sets) compared to verification signatures (individual references) shown as box and whisker plots. Random models were produced for A) GSE10334 PeriodovsContvV, B) GSE16134 PeriodovsContvV, C) GSE16134 PeriodovsContvall, and D) GSE10334 PeriodovsContvall.

3.4 Gene Set Enrichment Analysis Based Meta-analysis Verified Periodontitis Probes

Individual probes can be verified by examining leading-edge probes identified by GSEA across verification signatures (Table 2). This examination found 23 positive and 33 negative panel probes shared across all verification leading-edges regardless of dataset (STable 4). These probes were summarized in Tables 6 and 7. Heat maps of the differential expression for these probes across all 8 signatures revealed all probes were statistically significant (Figures 4 and 5), and therefore could have been identified via single gene analysis using T-score alone.

Probe	Symbol	Description
221036_s_at	APH1B	APH1B gamma secretase subunit
206398_s_at	CD19	CD19 molecule
209933_s_at	CD300A	CD300a molecule
216412_x_at	CKAP2	cytoskeleton associated protein 2
228427_at	FBXO16	F-box protein 16
204683_at	ICAM2	intercellular adhesion molecule 2
217084_at	IGHA1	immunoglobulin heavy constant alpha 1
234390_x_at	IGHV3-54	None
217145_at	IGK	immunoglobulin kappa locus
234366_x_at	IGLJ2	immunoglobulin lambda joining 2
216430_x_at	IGLV1-44	immunoglobulin lambda variable 1-44
234851_at	IGLV2-5	immunoglobulin lambda variable 2-5 (pseudogene)
1554208_at	MEI1	meiosis inhibitor 1
203435_s_at	MME	membrane metallo-endopeptidase
1558837_a_at	None	None
1563509_at	None	None
234415_x_at	None	None
204852_s_at	PTPN7	protein tyrosine phosphatase, non-receptor type 7
212699_at	SCAMP5	secretory carrier membrane protein 5
206049_at	SELP	selectin P (granule membrane protein 140kDa, antigen CD62)
1552542_s_at	TAGAP	T-cell activation RhoGTPase activating protein
205943_at	TDO2	tryptophan 2,3-dioxygenase
205922_at	VNN2	vanin 2

TABLE 6: Positive Periodontitis Panel Probes Shared in Verification Leading-edges.

Probe	Symbol	Description
220016_at	AHNAK	AHNAK nucleoprotein
206385_s_at	ANK3	ankyrin 3, node of Ranvier (ankyrin G)
238332_at	ANKRD29	ankyrin repeat domain 29
204671_s_at	ANKRD6	ankyrin repeat domain 6
213618_at	ARAP2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2
236534_at	BNIPL	BCL2/adenovirus E1B 19kD interacting protein like
218792_s_at	BSPRY	B-box and SPRY domain containing
205379_at	CBR3	carbonyl reductase 3
213050_at	COBL	cordon-bleu WH2 repeat protein
213379_at	COQ2	coenzyme Q2 4-hydroxybenzoate polyprenyltransferase
215103_at	CYP2C18	cytochrome P450, family 2, subfamily C, polypeptide 18
202428_x_at	DBI	diazepam binding inhibitor (GABA receptor modulator, acyl-CoA binding protein)
1558236_at	DR1	down-regulator of transcription 1, TBP-binding (negative cofactor 2)
203303_at	DYNLT3	dynein, light chain, Tctex-type 3
225846_at	ESRP1	epithelial splicing regulatory protein 1
223193_x_at	FAM162A	family with sequence similarity 162, member A
224345_x_at	FAM162A	family with sequence similarity 162, member A
218898_at	FAM57A	family with sequence similarity 57, member A
31874_at	GAS2L1	growth arrest-specific 2 like 1
232116_at	GRHL3	grainyhead-like 3 (Drosophila)
202967_at	GSTA4	glutathione S-transferase alpha 4
226188_at	LGALSL	lectin, galactoside-binding-like
225540_at	MAP2	microtubule-associated protein 2
203961_at	NEBL	nebulette
226490_at	NHSL1	NHS-like 1
231311_at	None	None
209570_s_at	NSG1	neuron specific gene family member 1
202289_s_at	TACC2	transforming, acidic coiled-coil containing protein 2
211382_s_at	TACC2	transforming, acidic coiled-coil containing protein 2
223694_at	TRIM7	tripartite motif containing 7
204141_at	TUBB2A	tubulin, beta 2A class IIa
204616_at	UCHL3	ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase)
1554628_at	ZNF57	zinc finger protein 57

TABLE 7: Negative Periodontitis Panel Probes Shared in Verification Leading-edges.

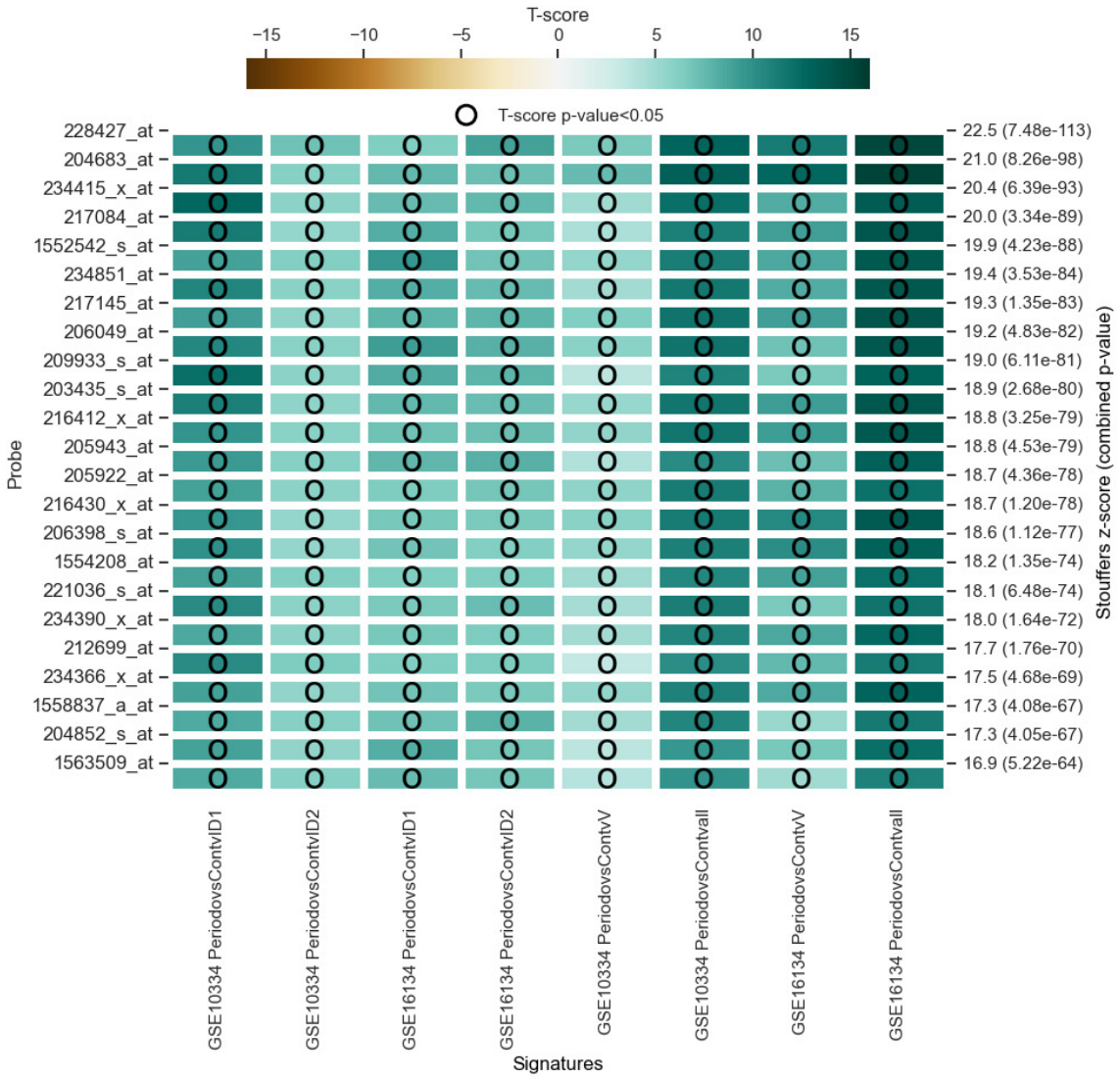

FIGURE 4: Heatmap of T-scores for shared positive periodontitis panel leading-edge probes.

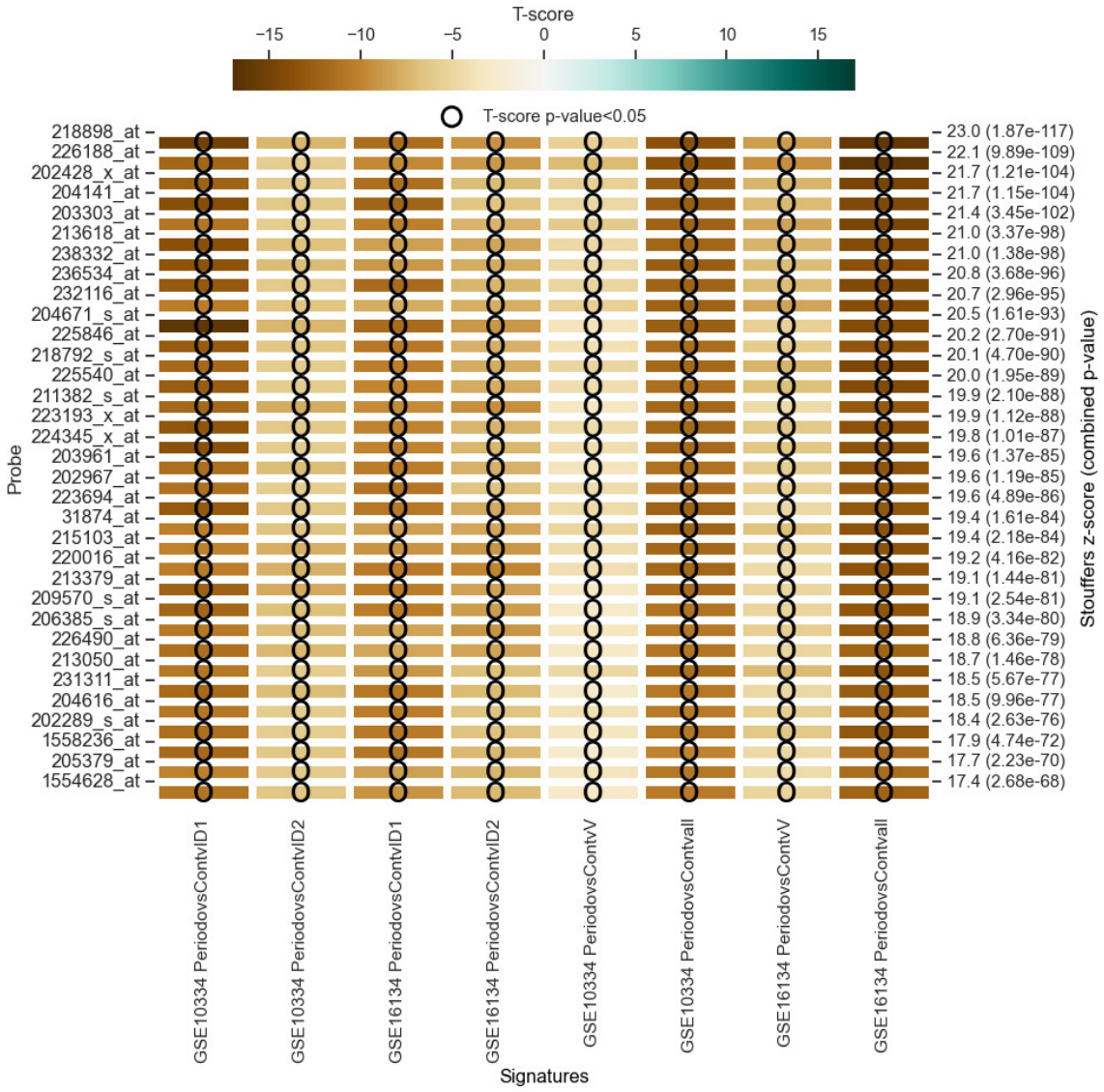

FIGURE 5: Heatmap of T-scores for shared negative periodontitis panel leading-edge probes.

4. DISCUSSION

Periodontitis is one of most common diseases associated with oral cavity and it can cause tooth loss and even impact general health of a patient. Despite efforts to develop effective treatments for Periodontitis, options remain limited, partially hindered by an incomplete comprehension of the molecular changes driving periodontitis. Identifying differentially expressed genes linked to periodontitis can enhance our insight into the molecular alterations induced by the condition, potentially contributing to the development of novel therapeutic strategies to combat periodontitis and address its future challenges. This study conducted a meta-analysis of gene signatures derived from mRNA expression data across various forms of periodontal infections to unveil differentially expressed genes associated with periodontitis.

There were 23 positive and 33 negative significantly differentially expressed genes in this study. There have been previous studies involving identifying genes associated with periodontitis and some genes identified in this study had a previous report indicating association with Periodontitis. CD19, a surface protein expressed on B cells, was identified to have association with

periodontitis in previous research (Hetta et al., 2020). ARAP2, which plays a complex role in regulating cell shape and movement by fine-tuning small protein signals inside the cell, have also been identified in previous research and this study to have association with Periodontitis (Kebschull et al., 2014). In addition, IGHA1, which plays a critical role in antibacterial immune response, was identified in previous research to have association with Periodontitis (Demmer et al., 2008). Finding common genes identified to have association with Periodontitis that have been discovered in previous studies reaffirms that the GSEA method was effective and accurate in identifying genes related to Periodontitis. Further, the GSEA method identified genes that had no prior connection with Periodontitis in previous studies that could be possibly targeted for treatment. For example, genes DYNLT3, SELP, and TDO2 did not have any prior connection to Periodontitis. This indicates that using the GSEA approach to find genes related to periodontitis enabled the identification of new genes that have not been previously identified. This will enable future research to possibly target genes identified in this study to develop treatment for Periodontitis. The genes identified in the study may lead to new diagnostic tests for early detection of periodontitis progression. For example, the expanded catalog of genes implicated in periodontitis could yield new prognostic biomarkers to predict disease severity and treatment outcomes (Jeon et al., 2020) On the therapeutic front, the results open up new possibilities for host-modulating therapies that correct aberrant inflammatory signaling pathways in periodontitis (Roky et al., 2020). Some of the genes found here could encode druggable enzymes like kinases, which small-molecule inhibitors could prevent or disable gum inflammation (Roky et al., 2020).

However, there are also a few limitations to this study. The datasets analyzed were limited to only two microarray studies, which may not fully capture the heterogeneity of gene expression across different populations and forms of periodontitis. Expanding the analysis to include additional datasets could improve reliability and generalizability of the results. Additionally, microarray technology is limited by potential cross-hybridization artifacts and inability to detect novel transcripts. Incorporating RNA-sequencing data in future analyses could provide higher resolution and more comprehensive characterization of the periodontitis transcriptome. Also, genes identified to have strong association with periodontitis have not been identified in the research. For example, genes IL6, IL10, and VDR have not been identified in this study (Tettamanti et al., 2017). Some genes related to Periodontitis might not have been identified because it didn't make top 500 t-score cut off despite there connection in previous studies. Also, heat map in figure 4 and 5 shows that all the genes identified to be associated with Periodontitis in this research had sufficient T-score in single gene analysis. This means that all the genes identified in GSEA meta analysis could have been identified in single gene analysis without being removed from the data because of its borderline significance. Although there was not any genes identified to be significant outside the cutoff of single gene analysis, the match between data of GSEA meta-analysis and single gene analysis further supports the conclusions that genes identified to be significant in this experiment were associated with Periodontitis and ruled out any possibility that there could have been any overlooked genes related to periodontitis due to its borderline significance in single gene analysis. An important next step is to validate differential expression of the identified genes using quantitative PCR in an independent cohort. It will also be critical to elucidate the functional roles of the novel gene candidates found to be associated with periodontitis, as they may represent promising therapeutic targets or biomarkers. However, GSEA in this study was sufficient to identify genes associated with Periodontitis by meta-analyzing over and under expressed genes in Periodontitis.

5. CONCLUSION

This study employed Gene Set Enrichment Analysis (GSEA) to meta-analyze gene signature in Periodontitis to locate any genes that have associations with Periodontitis. In this study, GSEA identified 23 significantly overexpressed and 33 significantly under expressed genes as being most associated with Periodontitis. Our research provides insights into gene expression associated with Periodontitis. By identifying these differentially expressed genes, future studies could potentially target these genes for the development of new diagnostic and therapeutic

strategies for this condition. Further research is needed to validate these findings and elucidate specific roles that these genes play in pathogenesis of periodontitis.

6. ACKNOWLEDGEMENTS

I would like to express my gratitude to Laura Harris for her contributions to this research manuscript. She provided assistance with writing manuscript, gene analysis on identified pathways, and graphic design.

7. REFERENCES

Abusleme, L., Hoare, A., Hong, B. Y., & Diaz, P. I. (2021). Microbial signatures of health, gingivitis, and periodontitis. *Periodontology 2000*, *86*(1), 57-78. https://doi.org/10.1111/prd.12362.

Cho, Y. D., Kim, W. J., Ryoo, H. M., Kim, H. G., Kim, K. H., Ku, Y., & Seol, Y. J. (2021). Current advances of epigenetics in periodontology from ENCODE project: a review and future perspectives. *Clinical Epigenetics*, *13*(1), 1-14. https://doi.org/10.1186/s13148-021-01074-w.

Gao, S., Lin, M., Chen, W., Chen, X., Tian, Z., Jia, T., ... & Wu, L. (2024). Identification of potential diagnostic biomarkers associated with periodontitis by comprehensive bioinformatics analysis. *Scientific Reports*, *14*(1), 93. https://doi.org/10.1038/s41598-023-50410-y.

Hajishengallis, G., &Korostoff, J. M. (2017). Revisiting the Page & Schroeder model: the good, the bad and the unknowns in the periodontal host response 40 years later. *Periodontology 2000*, *75*(1), 116-151. https://doi.org/10.1111/prd.12181.

Hetta, H. F., Mwafey, I. M., Batiha, G. E., Alomar, S. Y., Mohamed, N. A., Ibrahim, M. A., Elkady, A., Meshaal, A. K., Alrefai, H., Khodeer, D. M., & Zahran, A. M. (2020). CD19+ CD24hi CD38hi Regulatory B Cells and Memory B Cells in Periodontitis: Association with Pro-Inflammatory and Anti-Inflammatory Cytokines. *Vaccines*, *8*(2), 340. https://doi.org/10.3390/vaccines8020340.

Jeon, Y. S., Cha, J. K., Choi, S. H., Lee, J. H., & Lee, J. S. (2020). Transcriptomic profiles and their correlations in saliva and gingival tissue biopsy samples from periodontitis and healthy patients. *Journal of Periodontal & Implant Science*, *50*(5), 313. https://doi.org/10.5051/jpis.1905460273.

Jepsen, S., Caton, J. G., Albandar, J. M., Bissada, N. F., Bouchard, P., Cortellini, P., ... & Yamazaki, K. (2018). Periodontal manifestations of systemic diseases and developmental and acquired conditions: Consensus report of workgroup 3 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *Journal of clinical periodontology*, *45*, S219-S229. https://doi.org/10.1038/sj.bdj.2018.616.

Kebschull, M., Demmer, R. T., Grün, B., Guarnieri, P., Pavlidis, P., & Papapanou, P. N. (2014). Gingival tissue transcriptomes identify distinct periodontitis phenotypes. *Journal of dental research*, *93*(5), 459–468. https://doi.org/10.1177/0022034514527288.

Kim, Y. G., Kim, M., Kang, J. H., Kim, H. J., Park, J. W., Lee, J. M., ... & Lee, Y. (2016). Transcriptome sequencing of gingival biopsies from chronic periodontitis patients reveals novel gene expression and splicing patterns. *Human genomics*, *10*, 1-10. https://doi.org/10.1186/s40246-016-0084-0.

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, *2015*(11), pdb-top084970. https://doi.org/10.1101/pdb.top084970.

Kwon, T., Lamster, I. B., & Levin, L. (2021). Current concepts in the management of periodontitis. *International dental journal*, *71*(6), 462-476. https://doi.org/10.1111/idj.12630.

Li, Y., He, J., He, Z., Zhou, Y., Yuan, M., Xu, X., ... & Zhou, X. (2014). Phylogenetic and functional gene structure shifts of the oral microbiomes in periodontitis patients. *The ISME journal*, *8*(9), 1879-1891. https://doi.org/10.1038/ismej.2014.28.

Lundmark, A., Gerasimcik, N., Båge, T., Jemt, A., Mollbrink, A., Salmén, F., ... & Yucel-Lindberg, T. (2018). Gene expression profiling of periodontitis-affected gingival tissue by spatial transcriptomics. *Scientific reports*, *8*(1), 9370. https://doi.org/10.1038/s41598-018-27627-3.

Papapanou, P. N., Sanz, M., Buduneli, N., Dietrich, T., Feres, M., Fine, D. H., ... & Tonetti, M. S. (2018). Periodontitis: Consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions. *Journal of periodontology*, *89*, S173-S182. https://doi.org/10.1111/jcpe.12946.

Park, A., & Harris, L. K. (2021). Gene Expression Meta-Analysis Reveals Interferon-Induced Genes Associated With SARS Infection in Lungs. *Frontiers in immunology*, *12*, 694355. https://doi.org/10.3389/fimmu.2021.694355.

Peymani, F., Farzeen, A., & Prokisch, H. (2022). RNA sequencing role and application in clinical diagnostic. *Pediatric Investigation*, *6*(01), 29-35. https://doi.org/10.1002/ped4.12314.

Roky, M., Tan, J., Sztukowska, M. N., Trent, J. O., & Demuth, D. R. (2020). Identification of small-molecule inhibitors targeting Porphyromonasgingivalis interspecies adherence and determination of their in vitro and in vivo efficacies. *Antimicrobial Agents and Chemotherapy*, *64*(11), 10-1128. https://doi.org/10.1128/aac.00884-20

Sanz, M., Herrera, D., Kebschull, M., Chapple, I., Jepsen, S., Beglundh, T., Sculean, A., Tonetti, M. S., & EFP Workshop Participants and Methodological Consultants (2020). Treatment of stage I-III periodontitis-The EFP S3 level clinical practice guideline. Journal of clinical periodontology, 47 Suppl 22(Suppl 22), 4–60. https://doi.org/10.1111/jcpe.13290.

Sczepanik, F. S. C., Grossi, M. L., Casati, M., Goldberg, M., Glogauer, M., Fine, N., & Tenenbaum, H. C. (2020). Periodontitis is an inflammatory disease of oxidative stress: We should treat it that way. *Periodontology 2000*, *84*(1), 45-68. https://doi.org/10.1111/prd.12342.

Tettamanti, L., Gaudio, R. M., Iapichino, A., Mucchi, D., & Tagliabue, A. (2017). Genetic susceptibility and periodontal disease: a retrospective study on a large italian sample. *ORAL & implantology*, *10*(1), 20–27. https://doi.org/10.11138/orl/2017.10.1.020.

Tonetti, M. S., Greenwell, H., & Kornman, K. S. (2018). Staging and grading of periodontitis: Framework and proposal of a new classification and case definition. *Journal of periodontology*, *89*, S159-S172. https://doi.org/10.1002/jper.18-0006.

Trombelli, L., Farina, R., Silva, C. O., & Tatakis, D. N. (2018). Plaque-induced gingivitis: Case definition and diagnostic considerations. *Journal of clinical periodontology*, *45*, S44-S67. https://doi.org/10.1002/jper.17-0576.

Wang, X., Zou, Y., & Zhang, J. (2023). Identification of biomarkers related to prognosis and diagnosis of periodontitis by bioinformatics based on public database. *Oral Diseases*. https://doi.org/10.1111/odi.14740.